

REMARKS

Introductory Comments:

Claims 8, 9, 10, 11, 12, and 13 were examined in the Office Action under reply. Claims 8, 9, 10, 11, 12, and 13 stand rejected under 35 U.S.C. §101. Claims 9, 10, 11, 12, and 13 stand rejected under 35 U.S.C. §112, first paragraph. Claims 10, 11, 12 and 13 stand rejected under 35 U.S.C. §102(b) or (e) and claims 8, 9, 10, 11, 12, and 13 stand rejected under 35 U.S.C. §102(b) as anticipated. These rejections are believed to be overcome by the above amendments and are otherwise traversed for the reasons discussed below.

Overview of the Amendments:

Claims 1, 4-8 and 14-17 have been canceled as directed to non-elected subject matter. Cancellation of claims 1, 4-8 and 14-17 is without prejudice, without intent to abandon any originally claimed subject matter, and without intent to acquiesce in any rejections of the records. Applicants reserve the right to bring the canceled claim again in a related application.

Claims 9-13 have been amended to in order to recite the subject invention with greater particularity. Specifically, claims 9 and 11-13 have been amended to read on the elected polynucleotide SEQ ID NOs, make minor wording changes, and correct obvious typographical errors. Additionally, recitations from claims 10, 11, and 12 have been incorporated into claim 9.

New claims 18-24 have been added by this amendment. Support for claim 18, directed at a recombinant vector, is found in the specification at page 5, lines 7-8; and pages 8-28. Support for claim 19, directed at a host cell, is found in the specification at page 5, lines 7-8; and pages 8-28. Support for claim 20, directed at producing a recombinant polypeptide, is found in the specification at page 5, lines 22-24; and pages 8-28. Claim 21 corresponds to previous claim 9 with the non-elected sequences eliminated. Claims 22-24 correspond to claims 18-20 described above but depend from new claim 21.

Formal Matters:

Priority

The Examiner stated that priority has not been granted to the claimed international application PCT/IB98/01665 because no certified copy of the application was submitted to the Office. In lieu of submitting a costly certified copy of the 524 page application, applicants append hereto (1) a copy of Form PCT/IB308 "Notice Informing the Applicant of the

Communication of the International Application to the Designated Offices” and (2) a copy of the cover page of PCT/IB98/01665 indicating an October 9, 1998 filing date. In accordance with PCT Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s) (emphasis added). As indicated on the enclosed form PCT/IB308, notice was given to the U.S. by the International Bureau, thus no copy of the international application is required.

Sequencing Rules

The Examiner stated that the application failed to comply with the requirements of 37 CFR §§ 1.821-1.825 because certain sequences were not listed in the Sequence Listing. Applicants are providing a substitute specification that includes the information required by 37 CFR §§ 1.821-1.825. Specifically, applicants are providing a new Sequence Listing with all sequences disclosed in the filed specification. Additionally, the substitute specification includes sequence identifiers in the proper format at each sequence. In accordance with 37 CFR 1.821(f), the content of the sequence listing information recorded in computer readable form (submitted herewith) is identical to the written (paper) Sequence Listing (submitted herewith). The Sequence Listing includes no new matter.

Drawings

Applicants are submitting corrected drawings in accordance with Form PTO 948 under separate cover.

Objection to Title

The Examiner objected to the title, “Neisserial Antigens” as being not descriptive of the invention to which the claims are directed. In the attached substitute specification, the title of the application has been amended to “Neisserial Polynucleotides” to more clearly indicate the invention to which the elected claims are directed.

Brief Description of the Drawings

Applicants appreciate the Examiner pointing out the need for a recitation of subparts for various figures in the “Brief Description of the Drawings”. The substitute specification includes a recitation of subparts where appropriate.

Response to Claim Rejections

Claim Rejections under 35 U.S.C. §101

The Examiner rejected claims 8, 9, 10, 11, 12 and 13 under 35 U.S.C. §101, asserting “the claimed invention is directed to non-statutory subject matter” (Office Action, page 4). In order to facilitate prosecution, the term “isolated” has been added to claims 9, 10, 11, 12 and 13, as suggested by the Examiner. Accordingly, the rejections under 35 U.S.C. §101 should be withdrawn.

Claim Rejections under 35 U.S.C. §112, First Paragraph

The Examiner has rejected claims 9, 10, 11, 12 and 13 under 35 U.S.C. §112, first paragraph, asserting that the claims contain “subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention” (Office Action, page 5).

The Examiner argues: “given the broad scope of the claims due to the use of the open language ‘comprising’, they are drawn to a genus: any nucleotide that minimally contains the sequences of the claimed SEQ ID NOs, including full length genes, any fusion constructs, *etc.*” (Office Action, page 5).

Further, the Examiner argues that the:

“mere disclosure of a species: sequences of the elected SEQ ID NOs, does not provide adequate written description of the claimed genus. In view of the level of knowledge and skill in the art, one skilled in the art would not recognize from the disclosure that the applicant was in possession the genus of DNAs or RNAs encompassed in the claims which the sequences of the claimed SEQ ID NOs.” (Office Action, page 6).

However, applicants respectfully disagree.

In order to comply with the written description requirement, an applicant's specification must convey with reasonable clarity to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention, *i.e.*, whatever is now claimed. *Vas Cath Inc. v. Mahurkar*, 19 USPQ 1111, 1117 (Fed. Cir. 1991) (cited in MPEP §2163 and in the Examiner Guidelines on Written Description Requirement). The Examiner has the initial burden of presenting evidence or reasons why persons skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims. *In re Wertheim*, 191 USPQ 90 (CCPA 1976) (cited in MPEP §2163.04 in the Examiner Guidelines on Written Description Requirement). Moreover, it is axiomatic that a patent specification "need not teach, and preferably omits, what is well known in the art." See, *Spectra-Physics, Inc. v. Coherent, Inc.* 3 USPQ2d 1737, 1743 (Fed. Cir. 1987); *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 231 USPQ 81, 94 (Fed. Cir. 1986). Thus, determining whether the written description is satisfied requires reading the disclosure in light of the knowledge possessed by those skilled in the art. *In re Alton*, 37 USPQ2d 1578 (Fed. Cir. 1996).

The written description requirement does not necessitate the description of every species falling within the purview of a claimed genus. Further, satisfaction of the written description requirement does not require applicants to provide experimental data. Rather, the purpose of the written description requirement of 35 U.S.C. §112, first paragraph is to ensure that applicants were in possession of the claimed invention at the time of filing. *Vas Cath Inc. v. Mahurkar*, 19 USPQ 1111, 1117 (Fed. Cir. 1991) (cited in MPEP § 2163). Accordingly, the PTO Revised Examiner Guidelines on Written Description states:

Prior to determining whether the disclosure satisfies the written description requirement for the claimed subject matter, the examiner should review the claims and the entire specification, including the specific embodiments, figures, and sequence listings, to understand what applicant has identified as the essential distinguishing characteristics of the invention. ... *i.e.*, what the applicant has demonstrated possession of, and what applicant has claimed.

* * *

The written description requirement for a claimed genus may be satisfied through sufficient description of a representative number of species by actual reduction to practice...or by disclosure of relevant identifying characteristics, *i.e.*, structure or other physical

and/or chemical properties, by functional characteristics coupled with a known or disclosed correlation between function and structure, or by a combination of such identifying characteristics, sufficient to show the applicant was in possession of the claimed genus...A 'representative number of species means that the species which are adequately described are representative of the entire genus...Satisfactory disclosure of a 'representative number' depends on whether one of skill in the art would recognize that the applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus in view of the species disclosed (64 Fed. Reg. 71427, emphasis added)

The burden is on the Examiner to provide evidence as to why a skilled artisan would not have recognized that the applicants were in possession of claimed invention at the time of filing.

Applying these tenets, applicants submit that the Office has failed to carry its burden and that the present claims indeed comply with the written description requirement of 35 U.S.C. §112, first paragraph. A review of the application as a whole, coupled with the knowledge in the art at the time of filing, evidences that the application is more than sufficient to convey with reasonable clarity to those skilled in the art that, as of the filing date sought, they were in possession of the invention.

First, the written description requirement for a claimed genus may be satisfied through sufficient description of a representative number of species by actual reduction to practice. Contrary to the Office's position, applicants have indeed pointed to a number of specific DNA constructs and nucleic acid sequences (i.e., species) falling within the scope of the generic claims. Applicants call out specific DNA constructs, for example, polynucleotide open reading frames (ORFs) that were cloned into expression vectors such as pGEX, pTCR, pET, pGEX-His (Specification, page 69, lines 23-26, emphasis added). Indeed, "Table II – Summary of cloning, expression and purification" indicate that over 40 ORFs that contain the sequences of the claimed SEQ ID NOs were cloned into DNA constructs (Specification, Table II at pg 74-76). Accordingly, applicants have explicitly disclosed a large number of species falling within the generic claims.

Secondly, applicants have explicitly identified the essential distinguishing characteristics of the invention. At page 4, lines 14-15, applicants state the "invention provides nucleic acids comprising the Neisserial nucleotide sequences disclosed in the examples". The application

discloses 106 Examples and hundreds of polynucleotide sequences. Further, at page 5, lines 8-9 “the invention provides vectors comprising the nucleotide sequences of the invention (e.g. expression vectors)...”.

Based on the foregoing, there can be no doubt that applicants have identified a large number of species falling within the generic claims and identified the essential distinguishing characteristics of the invention. Therefore applicants have demonstrated possession of the claimed invention as set forth in the above Guidelines. Nevertheless, in order to hasten prosecution, applicants’ claims now include the recitation of an open reading frame. There is extensive support throughout the specification for ORFs containing the polynucleotides of the present invention including Table II at pg 74-76 and the 106 Examples. Hence, the Office’s rejection of the claims under 35 U.S.C. §112, first paragraph has been overcome and withdrawal thereof is respectfully requested.

The Examiner further rejected claims 12 and 13 under 35 U.S.C. §112, first paragraph, asserting that “the specification is not deemed to provide reasonable support to one of ordinary skill in the art that the biochemical activity of a nucleic acid at least 50% but less than 100% identical to the entire length of the elected sequences would be the same” (Office Action, page 6). Applicants respectfully disagree. Nevertheless, in order to advance prosecution, applicants have amended claim 12 to raise the percent identity to “90%.” Claim 9 as amended, also recites the percent identity of 90%, and claim 13 now depends from claim 9. Support for amended claims 9 and 12 is found in the specification at page 8, line 4.

In view of the above arguments and amendments, the applicants submit that the pending claims reasonably convey the claimed invention to one of ordinary skill in the art. Accordingly, the rejection of the claims under 35 U.S.C. §112, first paragraph first paragraph has been overcome and withdrawal thereof is respectfully requested.

Claim Rejections under 35 U.S.C. §102

Claims 10, 11, 12 and 13 were rejected under 35 U.S.C. §102(b) or (e) as anticipated by the various GenEmbl sequences or U.S. patents listed in the table on page 10 of the Office Action under reply. Specifically, the Examiner has rejected claim 10, and dependent claims 11, 12 and 13 asserting that the claims are anticipated by the disclosure of various database sequences and US patents that comprise a fragment of at least 10 base pairs of sequences of the elected

SEQ ID NOs, as required by claim 10 and as defined in the specification for the term “fragment” (page 4).

In order to facilitate prosecution, claim 10 has been amended to recite “a fragment greater than 18 nucleotides in length ...”. Support for amended claim 10 is found in the specification at page 4, lines 3-4. None of the database sequences or U.S. patents cited disclose a sequence greater than 18 nucleotides that is identical to SEQ ID NO: 125, SEQ ID NO: 127, SEQ ID NO: 131, SEQ ID NO: 463, SEQ ID NO: 465 SEQ ID NO: 569, or SEQ ID NO: 571. To further advance prosecution, SEQ ID NO: 651, SEQ ID NO: 649, SEQ ID NO: 653 have been deleted from claim 10.

In view of the above amendments and arguments, the cited reference sequences cannot be said to teach all the elements of the present invention. Accordingly, withdrawal of the above rejections is respectfully requested.

The Examiner also rejected claims 8-13 under 35 USC 102(b) as being anticipated by Paruchuri *et al.* (PNAS, USA, Vo. 87, No. 1, pages 333-3337, 1990). The Examiner states that Paruchuri *et al.* describes the isolation of chromosomal DNA from wild-type *Niesseria gonorrhea*. The Examiner further alleges: “since the nucleic acid sequences of the elected SEQ ID NOs are from *Niesseria*, it is inherent that the nucleic acid molecules i.e. the Neisserial chromosomal DNA, disclosed by Paruchuri *et al.* encode the proteins encoded by the nucleic acid sequences of the elected SEQ ID NOs, as required in claim 8 (Office Action, page 11). Additionally, the Examiner argues the DNA molecules disclosed in Paruchuri *et al.* comprise the nucleotide sequences of the elected SEQ ID NOs, and fragments thereof, as recited in claims 9 and 10. The Examiner further contends that DNA molecules disclosed in Paruchuri *et al.* inherently comprise the sequences complementary to the sequences of the elected SEQ ID NOs, or fragments thereof, as specified in claim 11. The Examiner further alleges the DNA molecules disclosed in Paruchuri *et al.* inherently comprise sequences that are at least 50% identical to the sequences of the elected SEQ ID NOs, or fragments thereof as recited in claim 12. Lastly, the Examiner argues that DNA molecules disclosed in Paruchuri *et al.* can inherently hybridize to the sequences of the elected SEQ ID NOs, or fragments thereof, or complements thereof, as specified in claim 13. Applicants respectfully disagree that Paruchuri *et al.* anticipates claims 9-13. The Examiner’s rejection of claim 8 is moot.

A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described in a single prior art reference. *Verdegaal Bros. v. Union Oil of California*, 814 F.2d 628, 631 (Fed.Cir. 1987); See also MPEP § 2131. As the Examiner correctly stated, Paruchuri *et al.* discloses the isolation of the chromosomal DNA from wild-type *Niesseria gonorrhea* (Office action at page 9). Paruchuri *et al.* did not determine the nucleotide sequence of any part of the *Niesseria gonorrhea* chromosome. There is nothing to suggest Paruchuri *et al.* contains any teachings of the identical chemical structure of the isolated polynucleotides in claims 9-13.

Paruchuri *et al.* does not isolate polynucleotides as claimed. Moreover, reliance on Paruchuri *et al.* is misplaced and applicants respectfully disagree. Inherency, may not be established by probabilities or possibilities. The mere fact that a certain characteristic may be present in the prior art is not sufficient to establish inherency of that characteristic. *Scaltech Inc. v. Retec/Tetra LLC* 156 F. 3d 1193 (Fed Cir. 1999). To establish inherency, the extrinsic evidence “must make clear that the missing descriptive matter is necessarily present in the thing described in the reference, and that it would be so recognized by persons of ordinary skill in the art.” *In re Robertson*, 49 USPQ2d 1949, 1950-51 (Fed. Cir. 1999), *quoting Continental Can v. Monsanto Co.*, 948 F.2d 1264, 1268, 20 USPQ2d 1746, 1749 (Fed. Cir. 1991).

Thus, the fact that polynucleotide sequences may be present in the chromosomal DNA from *Niesseria* is not sufficient to establish inherency of the claimed isolated polynucleotide sequences. The Examiner fails to establish inherency because Paruchuri *et al.* “must make clear that the missing descriptive matter [the isolated polynucleotide sequences as defined by the SEQ ID NOs] is necessarily present in the thing described in the reference [chromosomal DNA], and that it would be so recognized by persons of ordinary skill in the art.” The Examiner has provided no evidence that those of ordinary skill in the art would have recognized the chemical structure of the particular isolated polynucleotide sequences claimed in claims 9-13, from the mere disclosure of chromosomal DNA. Further, the Examiner’s reliance on *In re Best* and *In re Fitzgerald* is misplaced and applicants respectfully disagree because the Examiner “must provide a basis in fact and or/technical reasoning to reasonably support the determination that the alleged inherit characteristic necessarily flows from the teaching of the applied prior art”. *Ex parte Levy* 17 USPQ2d 1461 (Bd. Pat. App. & Inter. 1990) (emphasis in original). Here the Examiner has

failed to provide objective evidence or cogent technical reasoning to support the conclusion of inherency.

Thus, the Examiner's reliance on Paruchuri *et al.* is insufficient as a matter of law to meet the requirements of inherency. The Paruchuri *et al.* reference does not inherently disclose the isolated polynucleotides in claims 9-13. Accordingly, withdrawal of the rejections under 35 U.S.C. §102 (b) is respectfully requested.

Claim Objections

The Examiner objected to claims 8, 9 and dependent claims 11-13, asserting that these claims do not reflect the elected subject matter. Applicants elected 10 *polynucleotide* sequences or fragments thereof. Claim 8 depends from any one of the claims 4-6, which are drawn to *polypeptides*. Applicants have canceled claim 8 and amended claims 9, and 11-13 to read on the elected polynucleotide SEQ ID NOs as requested by the examiner. Accordingly, this basis for objection has been overcome.

The Examiner objected to claim 9 because it ends with two periods. Claim 9 has been amended to correct the typographical error. Thus, this basis for objection has also been overcome.

CONCLUSION

Applicants respectfully submit that the claims are novel and nonobvious and comply with the requirements of 35 U.S.C. 112. Accordingly, allowance is believed to be in order and an early notification to that effect is respectfully requested.

Please direct all further written communications regarding this application to

Alisa Harbin, Esq.
Intellectual Property –R440
P.O Box 8097
Emeryville, CA 94662-8097


Dated:

Chiron Corporation
ATTN: Intellectual Property -R440
P.O Box 8097
Emeryville, CA 94662-8097
Tel: (650) 843-5000 Fax: (650) 857-0663

Respectfully submitted,

COOLEY GODWARD LLP

By:



Roberta L. Robins

Reg. No. 33,208



NEISSERIAL [ANTIGENS] POLYNUCLEOTIDES

This application is a continuation-in-part of international patent application PCT/IB98/01665, filed October 9, 1998, from which priority is claimed under 35 U.S.C. § 119.

This invention relates to antigens from *Neisseria* bacteria.

5 BACKGROUND ART

Neisseria meningitidis and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely
10 different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In:
15 *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1
20 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis*
25 Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during

epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

- 5 Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a
- 10 tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the
- 15 vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).
- 20 Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and
- 25 therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular[?], *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different porins have been constructed (*eg.* Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (*eg.* Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (*eg.* EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (*eg.* 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the

Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

5 The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

10 The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

15 According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide sequences disclosed in the examples.

20 Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least *n* consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, *n* is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

25 According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various
5 forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term “nucleic acid” includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

10 According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*eg.* as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid,
15 protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B
20 or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

25 A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing
5 conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the
10 invention (*eg.* to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

The practice of the present invention will employ, unless otherwise indicated, conventional
15 techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *eg.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J.
20 Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology*
25 (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

5 Definitions

A composition containing X is “substantially free of” Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

10 The term “comprising” means “including” as well as “consisting” *eg.* a composition “comprising” X may consist exclusively of X or may include something additional to X, such as X+Y.

15 A “conserved” *Neisseria* amino acid fragment or protein is one that is present in a particular *Neisseria* protein in at least x% of *Neisseria*. The value of x may be 50% or more, e.g., 66%, 75%, 80%, 90%, 95% or even 100% (i.e. the amino acid is found in the protein in question in all *Neisseria*). In order to determine whether an amino acid is “conserved” in a particular *Neisseria* protein, it is necessary to compare that amino acid residue in the sequences of the protein in question from a plurality of different *Neisseria* (a reference population). The reference population may include a number of different *Neisseria* species or may include a single species. The reference population may include a number of different serogroups of a particular species or a single serogroup. A preferred reference population consists of the 5 most common *Neisseria*. The term
20 “heterologous” refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a *Neisseria* sequence is heterologous to a mouse host cell. A further examples would be two
25 epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An “origin of replication” is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain
5 origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A “mutant” sequence is defined as DNA, RNA or amino acid sequence differing from but having
10 sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an “allelic variant” of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid
15 molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5’ or 3’ untranslated regions of
20 the gene, such as in regulatory control regions (*eg.* see US patent 5,753,235).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

25 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3’) transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5’ end of the coding

sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element
5 determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences.
10 Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

15 The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped
20 orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus
25 [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen
5 bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved
10 either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells
15 are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem.*
20 *Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and
25 transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as

mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*eg.* Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the

heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired,

methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein
5 comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector
10 and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion
15 can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

20 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein,
25 which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells

infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence.

These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art.

5 Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 10 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant* 15 *Physiology*,. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. 25 Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A

general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed

out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

5 The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and 10 Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

15 The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

20 All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, 25 *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*,

Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue
5 is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop
10 simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the
15 invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be
20 adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation
25 region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits

negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences.

Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters.

For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid

promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that

preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an

extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia*

coli [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; Streptococcus lividans [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], Streptomyces lividans [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl₂ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, Campylobacter], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, Staphylococcus], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology I*:412, Streptococcus].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the

“TATA Box”) and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) “The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*,” in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with
5 cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two
10 amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific
15 processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites
20 encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US
25 patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor

fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (eg. see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences

flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*,
5 *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable
10 integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to
15 tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation
20 vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol.*
25 *Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.*

(1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse
5 (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*];
10 [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos.
15 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

Antibodies

20 As used herein, the term “antibody” refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An “antibody combining site” is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. “Antibody” includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised
25 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (*eg.* 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (*eg.* hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (*eg.* in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount

in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories; and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

10 Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with “pharmaceutically acceptable carriers,” which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents (“adjuvants”). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds.

Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) RibiTM adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (3) saponin adjuvants, such as StimulonTM (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (*eg.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*eg.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), *etc.*; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59TM are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral,

adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses *eg.* MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*eg.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly

preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or
5 collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698,
10 WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller
15 (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in
20 WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992)
25 *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5

native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC

VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example

ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex

beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

- 5 Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like
- 10 polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci.*
- 15 *USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033
- 20 Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.
- 25 A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects
5 can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or
10 transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic
15 cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well
20 known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

25 One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating

factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

5 B.Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, etc.

10 Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

15 The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim.*
20 *Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified
25 transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE
5 (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids
10 (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the
15 art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta*
20 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

25 E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally

occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

- 5 Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

- 10 A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

- The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 15 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

- Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of 20 naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

- Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. 25 Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for

example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

5 F.Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired
10 location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from
15 DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

20 The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when
25 combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods.

5 Antibodies to Neisserial proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody
10 or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed
15 by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

“Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen
20 bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt’s reagent or BLOTTO); concentration
25 of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

20 Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

25 The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be

complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe
5 can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases
10 or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the
15 complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.
20

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase
25 *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two “primer” nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37 (Fig. 1A-1E), 5 (Fig. 2A-2B), 2 (Fig. 3A-3D), 15 (Fig. 4A-4C), 22 (Fig. 5A-5C), 28 (Fig. 6A-6B), 32 (Fig. 7A-7B), 4 (Fig. 8A-8F), 61 (Fig. 9), 76 (Fig. 10A-10C), 89 (Fig. 11), 97 (Fig. 12A-12E), 106 (Fig. 13A-7C), 138 (Fig. 14A-B), 23 (Fig. 15A-15C), 25 (Fig. 16A-16E), 27 (Fig. 17A-17B), 79 (Fig. 18A-18B), 85 (Fig. 19A-19D) and 132 (Fig. 20A-20C). M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (◆) shows preimmune data; a triangle (▲)

shows GST control data; a circle () shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

Figure 21 shows an alignment comparison of amino acid sequences for ORF 4 for several strains of *Neisseria*. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. The Figure demonstrates a high degree of conservation among the various strains, further confirming its utility as an antigen for both vaccines and diagnostics.

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of

known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also
5 Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

10 To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID NO: 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID NO: 11)
15 represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical
20 evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORF

FINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psort.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

5 Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

10 The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

15 **A) Chromosomal DNA preparation**

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C
20 for 2 hours. Two phenol extractions (equilibrated to pH 8) and one ChCl_3 /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

25 **B) Oligonucleotide design**

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A

sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included a *Xho*I restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *Bam*HI-*Xho*I or *Eco*RI-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

10 5'-end primer tail: CGCGGATCCCATATG (SEQ ID NO: 1099) (*Bam*HI-*Nde*I)
 CGCGGATCCGCTAGC (SEQ ID NO: 1100) (*Bam*HI-*Nhe*I)
 CCGGAATTCTAGCTAGC (SEQ ID NO: 1101) (*Eco*RI-*Nhe*I)
 3'-end primer tail: CCCGCTCGAG (SEQ ID NO: 1102) (*Xho*I)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *Xho*I primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (SEQ ID NO: 1103) (*Nde*I)

5'-end primer tail: CGGGATCC (*Bam*HI)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *Nhe*I-*Bam*HI restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (SEQ ID NO: 1104) (*Nhe*I)

3'-end primer tail: CGGGATCC (*Bam*HI)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing

nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

- 5 The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

10 **Table I** shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, 15 once the complete sequence has been identified, this approach is generally no longer necessary.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG [<SEQ ID 924>] (SEQ ID NO: 924)	BamHI-NheI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT [<SEQ ID 925>] (SEQ ID NO: 925)	XhoI
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG [<SEQ ID 926>] (SEQ ID NO: 926)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG [<SEQ ID 927>] (SEQ ID NO: 927)	XhoI
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG [<SEQ ID 928>] (SEQ ID NO: 928)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGATTACGGACGCGCA [<SEQ ID 929>] (SEQ ID NO: 929)	XhoI
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC [<SEQ ID 930>] (SEQ ID NO: 930)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC [<SEQ ID 931>]	XhoI

ORF 5	Forward	(SEQ ID NO: 931) GGAATTCCATATGGCCATGG-TGGAAGGCGCACAAACC [<SEQ ID 932>] (SEQ ID NO: 932)	NdeI-NcoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAAC [<SEQ ID 933>] (SEQ ID NO: 933)	BamHI
	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG [<SEQ ID 934>] (SEQ ID NO: 934)	XhoI
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA [<SEQ ID 935>] (SEQ ID NO: 935)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC [<SEQ ID 936>] (SEQ ID NO: 936)	XhoI
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTC [<SEQ ID 937>] (SEQ ID NO: 937)	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTGCGGA [<SEQ ID 938>] (SEQ ID NO: 938)	XhoI
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC [<SEQ ID 939>] (SEQ ID NO: 939)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC [<SEQ ID 940>] (SEQ ID NO: 940)	XhoI
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGAAA [<SEQ ID 941>] (SEQ ID NO: 941)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCGAGGTTTTCGGG [<SEQ ID 942>] (SEQ ID NO: 942)	XhoI
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC [<SEQ ID 943>] (SEQ ID NO: 943)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAATGGGAAACCTTGTTTT [<SEQ ID 944>] (SEQ ID NO: 944)	XhoI
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG [<SEQ ID 945>] (SEQ ID NO: 945)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC [<SEQ ID 946>] (SEQ ID NO: 946)	XhoI
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC [<SEQ ID 947>] (SEQ ID NO: 947)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG [<SEQ ID 948>] (SEQ ID NO: 948)	XhoI
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG [<SEQ ID 949>] (SEQ ID NO: 949)	NdeI-NcoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG [<SEQ ID 950>] (SEQ ID NO: 950)	BamHI
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGCTATG [<SEQ ID 951>] (SEQ ID NO: 951)	XhoI
ORF 17	Forward	GGAATTCCATATGGCCATGG-TTGCCGGCCTGTTCG [<SEQ ID 952>] (SEQ ID NO: 952)	NdeI-NcoI

	Forward	CGGGATCC-ATTGCCGGCCTGTTCG [<SEQ ID 953>] (SEQ ID NO: 953)	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTTGTACAGC [<SEQ ID 954>] (SEQ ID NO: 954)	XhoI
ORF 18	Forward	GCGGATCCCATATG-ATTTTGCTGCATTTGGAT [<SEQ ID 955>] (SEQ ID NO: 955)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTTCTGAAAGC [<SEQ ID 956>] (SEQ ID NO: 956)	XhoI
ORF 19	Forward	GGAATTCATATGGCCATGG -TCGCCAGTGTTTTTACC [<SEQ ID 957>] (SEQ ID NO: 957)	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGTTTTTACCG [<SEQ ID 958>] (SEQ ID NO: 958)	BamHI
	Reverse	CCCGCTCGAG-GGTGTTTTTGAAGCTGCC [<SEQ ID 959>] (SEQ ID NO: 959)	XhoI
ORF 20	Forward	GGAATTCATATGGCCATGG -TCGGCGCGGGTATG [<SEQ ID 960>] (SEQ ID NO: 960)	NdeI-NcoI
	Forward	CGGGATCC-TTCGGCGCGGGTATG [<SEQ ID 961>] (SEQ ID NO: 961)	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGAGCA [<SEQ ID 962>] (SEQ ID NO: 962)	XhoI
ORF 22	Forward	GGAATTCATATGGCCATGG-TGATTAAAATCAAAAAGGTCT [<SEQ ID 963>] (SEQ ID NO: 963)	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAAAATCAAAAAGGTCTAAACC [<SEQ ID 964>] (SEQ ID NO: 964)	BamHI
	Reverse	CCCGCTCGAG-ATTATGATAGCGGCC [<SEQ ID 965>] (SEQ ID NO: 965)	XhoI
ORF 23	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC [<SEQ ID 966>] (SEQ ID NO: 966)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG [<SEQ ID 967>] (SEQ ID NO: 967)	XhoI
ORF 24	Forward	GGAATTCATATGGCCATGG -TGATGCCGAAATGGTG [<SEQ ID 968>] (SEQ ID NO: 968)	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGAAATGGTG [<SEQ ID 969>] (SEQ ID NO: 969)	BamHI
	Reverse	CCCGCTCGAG-TGTCAGCGTGCGCA [<SEQ ID 970>] (SEQ ID NO: 970)	XhoI
ORF 25	Forward	GCGGATCCCATATG-TATCGAAACTGATTGC [<SEQ ID 971>] (SEQ ID NO: 971)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG [<SEQ ID 972>] (SEQ ID NO: 972)	XhoI
ORF 26	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC [<SEQ ID 973>] (SEQ ID NO: 973)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGCGCGTTTT [<SEQ ID 974>] (SEQ ID NO: 974)	XhoI

ORF 27	Forward	GGAATTCCATATGGCCATGG-AGACCTATTCTGTTTA [<SEQ ID 974>] (SEQ ID NO: 1168)	NdeI-NcoI
	Forward	CGGGATCC- CAGACCTATTCTGTTTATTTTAATC [<SEQ ID 975>] (SEQ ID NO: 975)	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT [<SEQ ID 976>] (SEQ ID NO: 976)	XhoI
ORF 28	Forward	GGAATTCCATATGGCCATGG-ACGGCTGTACGTTGATGT [<SEQ ID 977>] (SEQ ID NO: 977)	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG [<SEQ ID 978>] (SEQ ID NO: 978)	BamHI
	Reverse	CCCGCTCGAG-TTTGTCTAGAGGAATTCGCG [<SEQ ID 979>] (SEQ ID NO: 979)	XhoI
ORF 29	Forward	GCGGATCCCATATG -AACGGTTTGGATGCCCCG [<SEQ ID 980>] (SEQ ID NO: 980)	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCCG [<SEQ ID 981>] (SEQ ID NO: 981)	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTGTCTAAGTTCCTGATATG [<SEQ ID 982>] (SEQ ID NO: 982)	XhoI
ORF 32	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG ([<SEQ ID 983>] (SEQ ID NO: 983)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTTTGTATGCTTTG [<SEQ ID 984>] (SEQ ID NO: 984)	XhoI
ORF 33	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG [<SEQ ID 985>] (SEQ ID NO: 985)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTTCAAACGGCC [<SEQ ID 986>] (SEQ ID NO: 986)	XhoI
ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT [<SEQ ID 987>] (SEQ ID NO: 987)	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT [<SEQ ID 988>] (SEQ ID NO: 988)	BamHI-NheI
	Reverse	CCCGCTCGAG-AAACAGCCATTTGAGCGA [<SEQ ID 989>] (SEQ ID NO: 989)	XhoI
ORF 37	Forward	GCGGATCCCATATG-GATGACGTATCGGATTTT [<SEQ ID 990>] (SEQ ID NO: 990)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAGCCCGCTTTCAGG [<SEQ ID 991>] (SEQ ID NO: 991)	XhoI
ORF 58	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT [<SEQ ID 992>] (SEQ ID NO: 992)	BamHI-NheI
	Reverse	CCCGCTCGAG-AGCATTGTCCAAGGGAC [<SEQ ID 993>] (SEQ ID NO: 993)	XhoI
ORF 65	Forward	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG [<SEQ ID 994>] (SEQ ID NO: 994)	NdeI-NcoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAGG [<SEQ ID	BamHI

ORF 66	Reverse	995>] (SEQ ID NO: 995) CCCGCTCGAG-CCGCATCGGCAGACA [<SEQ ID 996>] (SEQ ID NO: 996)	XhoI
	Forward	GCGGATCCCATATG-TACGCATTTACCGCCG [<SEQ ID 997>] (SEQ ID NO: 997)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGGATTTTGCAGAGATGG [<SEQ ID 998>] (SEQ ID NO: 998)	XhoI
ORF 72	Forward	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA [<SEQ ID 999>] (SEQ ID NO: 999)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA [<SEQ ID 1000>] (SEQ ID NO: 1000)	XhoI
ORF 73	Forward	GCGGATCCCATATG-AGATTTTTTCGGTATCGG [<SEQ ID 1001>] (SEQ ID NO: 1001)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCATCTTTTTCATGTTTCG [<SEQ ID 1002>] (SEQ ID NO: 1002)	XhoI
ORF 75	Forward	GCGGATCCCATATG- TCTGTCTTTCAAACGGC [<SEQ ID 1003>] (SEQ ID NO: 1003)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTTTTGCAAGACAG [<SEQ ID 1004>] (SEQ ID NO: 1004)	XhoI
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC [<SEQ ID 1005>] (SEQ ID NO: 1005)	NheI-NdeI
	Reverse	CGGGATCC-TTACGGTTTGACACCGTT [<SEQ ID 1006>] (SEQ ID NO: 1006)	BamHI
ORF 79	Forward	CGCGGATCCCATATG-GTTTCCGCCGCCG [<SEQ ID 1007>] (SEQ ID NO: 1007)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG [<SEQ ID 1008>] (SEQ ID NO: 1008)	XhoI
ORF 83	Forward	GCGGATCCCATATG-AAAACCCTGCTGCTGC [<SEQ ID 1009>] (SEQ ID NO: 1009)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCCGCCTTTGCGGC [<SEQ ID 1010>] (SEQ ID NO: 1010)	XhoI
ORF 84	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG [<SEQ ID 1011>] (SEQ ID NO: 1011)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTTTGCCGATCCGACCA [<SEQ ID 1012>] (SEQ ID NO: 1012)	XhoI
ORF 85	Forward	CGCGGATCCCATATG- GCGGTTTGGGGCGGA [<SEQ ID 1013>] (SEQ ID NO: 1013)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCGGCGCGGCGGGC [<SEQ ID 1014>] (SEQ ID NO: 1014)	XhoI
ORF 89	Forward	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA [<SEQ ID 1015>] (SEQ ID NO: 1015)	NdeI-NcoI
	Forward	CGGGATCC-GCCATACCTTCTTATCAGAG [<SEQ ID 1016>] (SEQ ID NO: 1016)	BamHI

ORF 97	Reverse	CCCGCTCGAG-TTTTTTGCATTAGAAAAAGC [<SEQ ID 1017>] (SEQ ID NO: 1017)	XhoI
	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC [<SEQ ID 1018>] (SEQ ID NO: 1018)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTTG [<SEQ ID 1019>] (SEQ ID NO: 1019)	XhoI
ORF 98	Forward	GCGGATCCCATATG-ACGGTAACTGCGG [<SEQ ID 1020>] (SEQ ID NO: 1020)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTGTTTCGGGCAAATC [<SEQ ID 1021>] (SEQ ID NO: 1021)	XhoI
ORF 100	Forward	GCGGATCCCATATG-TCGGGCATTTACACCG [<SEQ ID 1022>] (SEQ ID NO: 1022)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTTCGGCGGAA [<SEQ ID 1023>] (SEQ ID NO: 1023)	XhoI
ORF 101	Forward	GCGGATCCCATATG-ATTTATCAAAGAAACCTC [<SEQ ID 1024>] (SEQ ID NO: 1024)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTTCCGCCTTTCAATGT [<SEQ ID 1025>] (SEQ ID NO: 1025)	XhoI
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC [<SEQ ID 1026>] (SEQ ID NO: 1026)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAACACGAC [<SEQ ID 1027>] (SEQ ID NO: 1027)	XhoI
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC [<SEQ ID 1028>] (SEQ ID NO: 1028)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC [<SEQ ID 1029>] (SEQ ID NO: 1029)	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC [<SEQ ID 1030>] (SEQ ID NO: 1030)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC [<SEQ ID 1031>] (SEQ ID NO: 1031)	XhoI
ORF 105	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC [<SEQ ID 1032>] (SEQ ID NO: 1032)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG [<SEQ ID 1033>] (SEQ ID NO: 1033)	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG [<SEQ ID 1034>] (SEQ ID NO: 1034)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTCCCGATGATGTT [<SEQ ID 1035>] (SEQ ID NO: 1035)	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG [<SEQ ID 1036>] (SEQ ID NO: 1036)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCAACCGAAG [<SEQ ID 1037>] (SEQ ID NO: 1037)	XhoI

ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC [<SEQ ID 1038>] (SEQ ID NO: 1038)	EcoRI
	Reverse	AAACTGCAG-GGAAAACCATCCGCACTCTGCC [<SEQ ID 1039>] (SEQ ID NO: 1039)	PstI
ORF111	Forward	AAAGAATTC-GCACC GCAAAGGCAAAAACCGCA [<SEQ ID 1040>] (SEQ ID NO: 1040)	EcoRI
	Reverse	AAACTGCAG-TCTGCGCGTTTTCGGGCAGGGTGG [<SEQ ID 1041>] (SEQ ID NO: 1041)	PstI
ORF113	Forward	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAAC CG [<SEQ ID 1042>] (SEQ ID NO: 1042)	EcoRI
	Reverse	AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG [<SEQ ID 1043>] (SEQ ID NO: 1043)	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAAACAGAAAAGACGG [<SEQ ID 1044>] (SEQ ID NO: 1044)	EcoRI
	Reverse	AAAAAAGTCGAC-CTATTTT TAGGGGCTTTTGC TTGTTGAAAAGCCTGCC [<SEQ ID 1045>] (SEQ ID NO: 1045)	Sall
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG [<SEQ ID 1046>] (SEQ ID NO: 1046)	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTGGCC [<SEQ ID 1047>] (SEQ ID NO: 1047)	PstI
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG [<SEQ ID 1048>] (SEQ ID NO: 1048)	EcoRI
	Reverse	AAACTGCAG-CGGTTTGGCTGCCTGGCCGTTGAT [<SEQ ID 1049>] (SEQ ID NO: 1049)	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC [<SEQ ID 1050>] (SEQ ID NO: 1050)	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC [<SEQ ID 1051>] (SEQ ID NO: 1051)	PstI
ORF122	Forward	AAAAAAGTCGAC-ATGTCTTACCGCGCAAGCAGTTCTCC [<SEQ ID 1052>] (SEQ ID NO: 1052)	Sall
	Reverse	AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC [<SEQ ID 1053>] (SEQ ID NO: 1053)	PstI
ORF125	Forward	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT [<SEQ ID 1054>] (SEQ ID NO: 1054)	EcoRI
	Reverse	AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG [<SEQ ID 1055>] (SEQ ID NO: 1055)	PstI
ORF126	Forward	AAAGAATTC-GCGGAAACGGTCAAG [<SEQ ID 1056>] (SEQ ID NO: 1056)	EcoRI
	Reverse	AAACTGCAG-TTAATCTTGTCTTCCGATATAC [<SEQ ID 1057>] (SEQ ID NO: 1057)	PstI
ORF127	Forward	AAAGAATTC-ATGACTGATAATCGGGGTTTACG [<SEQ ID 1058>] (SEQ ID NO: 1058)	EcoRI

ORF128	Reverse	AAAAAAGTCGAC-CTTAAGTAACTGTCAGTCCTTATC [<SEQ ID 1059>] (SEQ ID NO: 1059)	SalI
	Forward	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC [<SEQ ID 1060>] (SEQ ID NO: 1060)	EcoRI
	Reverse	AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTTGAGCAGGC G [<SEQ ID 1061>] (SEQ ID NO: 1061)	PstI
ORF129	Forward	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATAC CG [<SEQ ID 1062>] (SEQ ID NO: 1062)	EcoRI
	Reverse	AAACTGCAG-TTATTTTTTTGATGAAATTTTGGGGCGG [<SEQ ID 1063>] (SEQ ID NO: 1063)	PstI
ORF130	Forward	AAAGAATTC-GCAGTACTTGCCATTCTCGGTGCG [<SEQ ID 1064>] (SEQ ID NO: 1064)	EcoRI
	Reverse	AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT [<SEQ ID 1065>] (SEQ ID NO: 1065)	PstI
ORF 131	Forward	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT [<SEQ ID 1066>] (SEQ ID NO: 1066)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAGCGGACGCGTTC [<SEQ ID 1067>] (SEQ ID NO: 1067)	XhoI
ORF 132	Forward	GCGGATCCCATATG-AAAGAAGCGGGGTTTG [<SEQ ID 1068>] (SEQ ID NO: 1068)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAATCTGCCAGCCGT [<SEQ ID 1069>] (SEQ ID NO: 1069)	XhoI
ORF 133	Forward	CGCGGATCCCATATG-GAAGATGCAGGGCGCG [<SEQ ID 1070>] (SEQ ID NO: 1070)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACTTGTAGCTCATCGT [<SEQ ID 1071>] (SEQ ID NO: 1071)	XhoI
ORF 134	Forward	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG [<SEQ ID 1072>] (SEQ ID NO: 1072)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCCTGTGCCAATGCG [<SEQ ID 1073>] (SEQ ID NO: 1073)	XhoI
ORF 135	Forward	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT [<SEQ ID 1074>] (SEQ ID NO: 1074)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAATACCGCTGAGGATG [<SEQ ID 1075>] (SEQ ID NO: 1075)	XhoI
ORF 136	Forward	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC [<SEQ ID 1076>] (SEQ ID NO: 1076)	BamHI-NheI
	Reverse	CCCGCTCGAG-TTCCGAATATTTGGAACTTTT [<SEQ ID 1077>] (SEQ ID NO: 1077)	XhoI
ORF 137	Forward	CGCGGATCCCATATG-GGCACGGCGGGAAATA [<SEQ ID 1078>] (SEQ ID NO: 1078)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAACGGTATGCCGCC [<SEQ ID 1079>] (SEQ ID NO: 1079)	XhoI

ORF 138	Forward	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC [<SEQ ID 1080>] (SEQ ID NO: 1080)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CGGCGTTTTATAGCGG [<SEQ ID 1081>] (SEQ ID NO: 1081)	XhoI
ORF 139	Forward	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG [<SEQ ID 1082>] (SEQ ID NO: 1082)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAACGTTTCCGTGCGTTT [<SEQ ID 1083>] (SEQ ID NO: 1083)	XhoI
ORF 140	Forward	GCGGATCCCATATG-TTGCCACAGGCAGC [<SEQ ID 1084>] (SEQ ID NO: 1084)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACGATGGCAAACAGC [<SEQ ID 1085>] (SEQ ID NO: 1085)	XhoI
ORF 141	Forward	GCGGATCCCATATG-CCGTCTGAAGCAGTCT [<SEQ ID 1086>] (SEQ ID NO: 1086)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCTGTTGTTTTTAAATATT [<SEQ ID 1087>] (SEQ ID NO: 1087)	XhoI
ORF 142	Forward	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG [<SEQ ID 1088>] (SEQ ID NO: 1088)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGTATAGCCTACCT [<SEQ ID 1089>] (SEQ ID NO: 1089)	XhoI
ORF 143	Forward	GCGGATCCCATATG-GATACCGCTTTGAACCT [<SEQ ID 1090>] (SEQ ID NO: 1090)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AATGGCTTCCGCAATATG [<SEQ ID 1091>] (SEQ ID NO: 1091)	XhoI
ORF 144	Forward	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC [<SEQ ID 1092>] (SEQ ID NO: 1092)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGATTGTTGTTGTTTTTCG [<SEQ ID 1093>] (SEQ ID NO: 1093)	XhoI
ORF 147	Forward	GCGGATCCCATATG-TCTGTCTTTCAAACGGC [<SEQ ID 1094>] (SEQ ID NO: 1094)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTTTTGCAAGACAG [<SEQ ID 1095>] (SEQ ID NO: 1095)	XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH_4OH , and deprotected by 5 hours incubation at 56°C . The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100 μl or 1ml of water. OD_{260} was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/ μl .

C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40 μM of each oligo, 400-800 μM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl_2), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10 μl DMSO or 50 μl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C .

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds $50-55^\circ\text{C}$	30-60 seconds 72°C
Last 30 cycles	30 seconds 95°C	30 seconds $65-70^\circ\text{C}$	30-60 seconds 72°C

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

D) Digestion of PCR fragments

- 10 The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:
- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
 - *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
 - For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
 - *EcoRI/PstI*, *EcoRI/Sall*, *Sall/PstI* for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion
- 20 Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by
- 25 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the
10 vector pTRC99 (Pharmacia).

F) Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the
15 manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boehringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3
20 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml
25 ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and

the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SalI* or, for ORF 122, *SalI-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C)

to OD₅₅₀ 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again.

- 5 The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia) (previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion
10 protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD₂₈₀ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must
15 be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis (ORFs 111-129)

- To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a
20 Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂ PO₄] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] overnight
25 at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal
5 temperature (20-37°C) to OD₅₅₀ 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

10 The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40
15 minutes.

Supernatants were collected and mixed with 150µl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a
20 disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D₂₈₀ of 0.02-0.06. The His-fusion
25 protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl,

100mM phosphate buffer, pH 4.5) and fractions collected until the $O.D_{280}$ was 0.1. 21 μ l of each fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20 μ g/ml using
5 dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

10
$$\text{Protein (mg/ml)} = (1.55 \times OD_{280}) - (0.76 \times OD_{260})$$

L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on
15 the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

20 M) Mice immunisations

20 μ g of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with $Al(OH)_3$ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice
25 were immunised using Freund's adjuvant, rather than $Al(OH)_3$, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56.

Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA was considered positive when OD₄₉₀ was 2.5 times the respective pre-immune sera.

O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were

let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

15 P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

R) Western blotting

- 5 Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was
10 washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

15 S) Bactericidal assay

- MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₆₂₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was
20 washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₆₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

- 50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well.
25 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on

Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II gives a summary of the cloning, expression and purification results.

TABLE II – Summary of cloning, expression and purification

5

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion

orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	
orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	

orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

Example 1

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 1>] (SEQ ID NO: 1):

```

5      1  ATGAAACAGA CAGTCAA .AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
      51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
     101  A .GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
     151  TAT .TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
     201  GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
    10  251  GCTGGATGTA TGCCAACGGG CGCGC .GTGC GCCAAGATGA TACCGAAGCG
     301  GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
     351  CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
     401  TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
     451  GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
    15  501  AGACCG . . .

```

This corresponds to the amino acid sequence [<SEQ ID 2; ORF37>] (SEQ ID NO: 2; ORF37):

```

      1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAAQGN AAAQYNLGAM
     51  YXQTRVRRD DAEAVRWYRQ PAEQGLAQAG YNLGWMYANG RXVRQDDTEA
    101  VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAG
    151  AQNNLGVMYA ERXRVRQD . . .
20

```

Further work revealed the complete nucleotide sequence [<SEQ ID 3>] (SEQ ID NO: 3):

```

      1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
     51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
    25  101  AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
     151  TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
     201  GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
     251  GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
     301  GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
    30  351  CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
     401  TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
     451  GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
     501  AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
     551  ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA
    35

```

This corresponds to the amino acid sequence [<SEQ ID 4; ORF37-1>] (SEQ ID NO: 4; ORF37-1):

```

      1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
     51  YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAG YNLGWMYANG RGVRQDDTEA
    101  VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAG
    151  AQNNLGVMYA ERRGVRQDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY*
40

```

Further work identified the corresponding gene in strain A of *N.meningitidis* [<SEQ ID 5>] (SEQ ID NO: 5):

```

5      1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
      51  GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
     101  AGGCGGCAGC ACAGGGAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
     151  TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
     201  GCTTGGCAAG GCTTGTCAAA ACGGATACCA AGACAGCTGC GACAATGACC
     251  AACGCCTGAA AGCGGGTTAT TGA

```

10 This encodes a protein having amino acid sequence [<SEQ ID 6; ORF37a>] (SEQ ID NO: 6; ORF37a):

```

      1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
     51  YAERRGVRQD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *

```

15 The originally-identified partial strain B sequence (ORF37) (SEQ ID NO: 2) shows 68.0% identity over a 75aa overlap with ORF37a (SEQ ID NO: 6):

```

      10      20      30      40      50      60
orf37.pep  MKQTVXMLAAALIALGLNRPVWXDDVSDFRNLXAAAQGNAAQYNLGAMYXQRTVRVRD
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf37a     MKQTVKWLAAALIALGLNQA VVWADDVSDFRNLQAAAQGNAAQNNLGVMYAERRGVRQD
      10      20      30      40      50      60

      70      80      90     100     110     120
orf37.pep  DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
      | | : | : : : |
orf37a     RALAQEWLGKACQNGYQDSCDNDQRLKAGYX
      70      80      90

```

Further work identified the corresponding gene in *N.gonorrhoeae* [<SEQ ID 7>] (SEQ ID NO: 7):

```

30      1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
      51  GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAAACTTGC
     101  AGgcggcagA ACaggGAAAT GCAGCAGCCC AATTCAATTT GGGCGTGATG
     151  TATGAAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
     201  GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTTGG
     251  GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TGCCTCGCT
     301  CAACAATGGC TTGGCAAGGC TTGTCAAAAC GGAGACCAAA ACAGCTGCGA
35     351  CAATGACCAA CGCCTGAAGG CGGGTTATTA A

```

This encodes a protein having amino acid sequence [<SEQ ID 8; ORF37ng>] (SEQ ID NO: 8; ORF37ng):

```

40      1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
      51  YENGQGVQRD YVQAVQWYRK ASEQGDAAQ YNLGLMYDYG RGVQRDLALA
     101  QQWLKGACQN GDQNSCDNDQ RLKAGY*

```

The originally-identified partial strain B sequence (ORF37) (SEQ ID NO: 2) shows 64.9% identity over a 111aa overlap with ORF37ng (SEQ ID NO: 8):

5	orf37.pep	MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAQGNAAQYNLGAMYXQRTVRVD	60
	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAQFNLGVMYENGQGVQRD	60
10	orf37.pep	DAEAVRWYRQPAEQGLAAQYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG	120
	orf37ng	YVQAVQWYRKASEQGDAAQYNLGLMYDGRGVRQDLALAQQWLGKACQNGDQNSCDNDQ	120
10	orf37.pep	VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERXVRQD	168
	orf37ng	RLKAGY	126

The complete strain B sequence (ORF37-1) (SEQ ID NO: 4) and ORF37ng (SEQ ID NO: 8) show 51.5% identity in 198 aa overlap:

15	orf37-1.pep	MKQTVKWLAAALIALGLNRAVWADDVSDFRENLQAAQGNAAQYNLGAMYKGRGVRRD	10 20 30 40 50 60
	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAQFNLGVMYENGQGVQRD	10 20 30 40 50 60
20	orf37-1.pep	DAEAVRWYRQAAEQGLAAQYNLGWMYANGRGVRQDDTEAVRWYRQAAAQGVVQAQYNLG	70 80 90 100 110 120
	orf37ng	YVQAVQWYRKASEQGDAAQYNLGLMYDGRGVRQD-----	70 80 90
25	orf37-1.pep	VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERRGVRRQDRALAEWFGKAC	130 140 150 160 170 180
	orf37ng	-----LALAQQWLGKAC	100
30	orf37-1.pep	QNGDQDGCNDQRLKAGYX	190 199
	orf37ng	QNGDQNSCDNDQRLKAGYX	110 120

Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (SEQ ID NO: 4) (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure

Example 2

10 TTCGCGCA CATCGCGCGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA
 GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGACTTGACC CGAAATCCTA
 TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG
 ACGTTTCCGC GCAAATCCTG ACTTCsGGAC TTTTGGGCGA GCAGTACATC
 GGGCTGCAGC AGGGCGGCGA CACGGAAAAC CTTGCTGCCG GCGACACCAT
15 CTCCGTAAC AGTTCTGCAA TGGTTCTGGA AAACCTTATC GGCAAATTCA
 TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGGCAATGC GGAAAAAGCC
 GCCGAATAA

20 1 FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD
 51 VSAQILTSGL LGEQYIGLQQ GGDTENLAAG DTISVTSSAM VLENLIGKFM
 101 TSFAEKNADG GNAEKAEE*

25 Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029 (SEQ
ID NO: 1105))

```

      20      30      40      50      60      70
yrbd.h LGIGALVFLGLRVANVQGFAETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAILDE
      |::| |||||::|::| :|::| |||::|::|
N.m      FGDIGGLKVNAPVKVSAGVLVGRVGAIGLDP
      10      20      30

```


5

```

      80      90      100      110      120      130
yrbd.h KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
      ||| ::|::: :| :::: | | |||||:::| | ||: | :|: |
N.m    KSYQARVRLDLGKY-QFSSDVSQAILTSGLLGEQYIGLQGG---GDTENLAAGDTISVT
      40      50      60      70      80

      140      150      160
yrbd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
      :|||:|:|:|: :::|:|:|: ::::|:
N.m    SSAMVLENLIGKFMTSFAEKNADGGNAEKAEX
      90      100      110      120
```

10

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID NO: 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae* (SEQ ID NO: 1106 yrbx):

15 yrbd GAAAVAF²⁰LA³⁰FRVAGGA⁴⁰AFGGSDKTYAVYAD⁵⁰FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
N.m FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
10 20 30

20 yrbd KSYQARVRLDLDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
N.m KSYQARVRLDLDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
40 50 60 70 80 90

25 yrbd VLENLIGKFMTSFAEKN¹⁴⁰AE¹⁵⁰GGNAE¹⁶⁰KAAEX
N.m VLENLIGKFMTSFAEKNADGGNAE¹⁰⁰KAAEX
100 110 120

30 The complete yrbd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

Example 3

35 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 11>] (SEQ ID
NO: 11):

40

1	..ATTTTGATAT	ACCTCATCCG	CAAGAACTCTA	GGTTCGCCCG	TCTTCTTCTT
51	TCAGGAAACGC	CCCGGAAAGG	ACGGAACC	TTTTAAATG	GTCAAATTCC
101	GTTCCATGCG	CGACGGCTTG	TATTCAGACG	GCATTCCGCT	GCCCGACGGA
151	GAACGCCTGA	CACCGTTCGG	CAAAAACTG	CGTGCCCGCA	GTWtGGACGA

5
10
15

```

201  ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
251  CCGCGCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAAC
301  CCGCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG
351  GCGCAACGCG CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
401  TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
451  AAAAAAGTAT TAATCAAGGA AGGGATTTCG GCACAGGGCG AACA.aCCAT
501  GCCCCCTTTC ACAGGAAAAC GCAAACCTCGC CGTCGTCGGT GCGGGCGGAC
551  ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA
601  ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
651  CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
701  ACGTCGCCGT CGCCGTCGGC AACAACCGCA TCCGCCGCCA AATCGCCGAA
751  AAAGCCGCCG CGCTCGGCTT CGCCCTGCCG GTACTGGTTC ATCCGGACGC
801  GACCGTCTCG CTTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
851  AAGCGGTCG.

```

This corresponds to the amino acid sequence [<SEQ ID 12; ORF3>] (SEQ ID NO: 12; ORF3):

20

```

1  .ILIIYLIRKLNL GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG
51  ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN
101 RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTV
151 KKVLIKEGIS AQGEXTMPFF TGKRKLAVVG AGGHGKVVD LAAALGRYRE
201 IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

```

25 Further sequence analysis revealed the complete nucleotide sequence [<SEQ ID 13>] (SEQ ID NO: 13):

30
35
40
45
50

```

1  ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
51  ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
101 AGAATCTAGG TTCGCCCCTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
151 GGAAAACCTT TTTAAATGGT CAAATTCCGT TCCATGCGCG ACGCGCTTGA
201 TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
251 AAAAACTGCG TGCCGCCAGT TTGACGAAC TGCTGAATT ATGGAATATC
301 TTTAAAGGCG AGATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA
351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAAACCG
401 GCATTACCGG CTGGCGCGAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
451 GAAAAATTCT CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT
501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
551 GGATTTCGCG ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
601 AAACTCGCCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT
651 TGCCGCCGCA CTCGGCCGGT ACAGGGAAAT CGTTTTTCTG GACGACCGCG
701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
751 GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
851 CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001 ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCTG
1051 GGCAACACGC ATATCGCGCA AGAAAGCTGG ATAGGCACGG GCGCGTCGAG
1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
1201 AAGCCGCTGC CGCGCAAAAA CCCCAGAGACC TCGACAGCAT AA

```

This corresponds to the amino acid sequence [<SEQ ID 14; ORF3-1>] (SEQ ID NO: 14; ORF3-1):

1 MSKFFKRLFD IVASASGLIF LSPVFLILIIY LIRKNL GSPV FFFQERP GKD

51	GKPFKMKVFR	SMRDALDSDG	IPLPDGERLT	PFGKKLRAAS	LDELPELWNI
101	LKGEMSLVGP	RPLLMQYLPL	YDNFQNRHE	MKPGITGWAQ	VNGRNALSWD
151	EKFACDVWYI	<u>DHFSLCDLIK</u>	<u>ILLTVKKVL</u>	IKEGISAQGE	ATMPPTGKR
201	KLAVVGAGGH	GKVVADLAAA	LGRYREIVFL	DDRAQGSVNG	FSVIGTTLLL
251	ENSLSPQYD	VAVAVGNMRI	RRQIAEKAAA	LGFAFPVLVH	PDATVSPSAT
301	VGQGSVVMKA	AVVQAGSVLK	DGVIVNTAAT	VDHDCLLNAP	VHISPGAHLN
351	INTHIGEESE	IGTGACSRQQ	IRIGSRATIG	AGAVVVRDVS	DGMTVAGNPA
401	KPLPRKNPET	STA*			

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 (SEQ ID NO: 12) shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) (SEQ ID NO: 16) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF3a nucleotide sequence [<SEQ ID 15>] (SEQ ID NO: 15) is:

45 1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG

5 51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
 101 AGAATCTGGG TTCGCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
 151 GGAAAACCTT TTAAATGGT CAAATTCCGT TCCATGCACG ACGCGCTTGA
 201 TTCAGACGGC ATTCTGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
 251 AAAAAGTGGC TGCCGCCAGT TTGGACGAAC TGCCCGAACT GTGGAACGTC
 301 CTCAAAGGCG ACATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAAACCGG
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
 451 GAACGCTTCG CATCGGACAT CTGGTATATC GACCACTTCA GCCTGTGCCT
 10 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAAGAAG
 551 GGATTTCGCG ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
 601 AAAGTTGCCG TCGTCGGTGC GGGCGGACAC GGCAAAGTCG TTGCCGAGCT
 651 TGCCGCCGCA CTCGGCACAT ACGGCGAAAT CGTTTTTCTG GACGACCGCG
 701 TCCAAGGCAG CGTCAACGGC TTCCCCGTCA TCGGCACGAC GCTGCTGCTT
 15 751 GAAAACAGTT TATCGCCCGA ACAATTCGAC ATCGCCGTCG CCGTCGGCAA
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCCGCGCG CTCGGCTTCG
 851 CCCTGCCCGT CCTGATTCAT CCGGACTCGA CCGTCTCGCC TTCTGCAACA
 901 GTCGGACAAG GCGGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCTGACAG
 951 CGTATTGAAA GACGGCGTAA TTGTGAACAC TGCCGCCACC GTCGATCACG
 20 1001 ATTGCCTGCT TGATGCTTTC GTCCACATCA GCCCGGGCGC GCACCTGTCTG
 1051 GGCAACACGC GTATCGGCGA AGAAAGCTGG ATAGGCACAG GCGCGTGCAG
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
 1151 TCGTCTGTGC CGACGTTTCA GACGGCATGA CCGTCGCGGG CAACCCGGCA
 25 1201 AAACCATTGG CAGGCAAAAA TACCGAGACC CTGCGGTCTG AA

This is predicted to encode a protein having amino acid sequence[<SEQ ID 16>] (SEQ ID NO: 16):

30 1 MSKFFKRLFD IVASASGLIF LSPVFLILY LIRKNLGSPV FFFQERPGKD
 51 GKPFKMKVFR SMHDALDSGD ILLPDGERLT PFGKKLRAAS LDELPELWNV
 101 LKGDMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNERNALSWD
 151 ERFACDIWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPTGKR
 201 KLAVVGAGGH GKVVAELAAA LGTYGEIVFL DDRVQGSVNG FPIVIGTLLL
 251 ENSLSPEQFD IAVAVGNRI RRQIAEKAAA LGFALPVLH PDSTVSPSAT
 301 VGQGGVMAK AVVQADSVLK DGVIVNTAAT VDHDCLLDAF VHISPGAHLS
 35 351 GNTRIGESW IGTGACSRQQ IRIGSRATIG AGAVVRDVS DGMTVAGNPA
 401 KPLAGKNTET LRS*

Two transmembrane domains are underlined.

ORF3-1 (SEQ ID NO: 14) shows 94.6% identity in 410 aa overlap with ORF3a (SEQ ID NO: 16):

40 10 20 30 40 50 60
 orf3a.pep MSKFFKRLFDIVASASGLIFLSPVFLILY LIRKNLGSPVFFQERPGKDGKPFKMKVFR
 orf3-1 MSKFFKRLFDIVASASGLIFLSPVFLILY LIRKNLGSPVFFQERPGKDGKPFKMKVFR
 10 20 30 40 50 60
 45 70 80 90 100 110 120
 orf3a.pep SMHDALDSGDILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL
 orf3-1 SMRDALDSGDIPDPGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL
 70 80 90 100 110 120
 50 130 140 150 160 170 180
 orf3a.pep YDNFQNRHEMKPGITGWAQVNERNALSWDERFACDIWYIDHFSCLDIKILLTVKKVL

	orf3-1	YDNFQNRRRHEMKPGITGWAQVNGRNLASWDEKFCADVWYIDHFSLCLDIKILLTVKKVL	130	140	150	160	170	180
5	orf3a.pep	IKEGISAQGEATMPFFTGKRKLAVVGAGGHGKVVAAELAAALGTYGEIVFLDDRVOGQSVNG	190	200	210	220	230	240
	orf3-1	IKEGISAQGEATMPFFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG	190	200	210	220	230	240
10	orf3a.pep	FPVIGTTLLLENSLSPEQFDIAVAVGNRRIRRQIAEKAAALGFALPVLHPDSTVSPSAT	250	260	270	280	290	300
	orf3-1	FSVIGTTLLLENSLSPEQYDVAVAVGNRRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	250	260	270	280	290	300
15	orf3a.pep	VGQGGVVMKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPAGHLSGNTRIGEEESW	310	320	330	340	350	360
	orf3-1	VGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPAGHLSGNTHIGEEESW	310	320	330	340	350	360
20	orf3a.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLAGKNTETLR SX	370	380	390	400	410	
	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX	370	380	390	400	410	

Homology with hypothetical protein encoded by yvfc gene (accession Z71928) (SEQ ID NO: 1108) of B. subtilis

ORF3 (SEQ ID NO: 12) and YVFC proteins (SEQ ID NO: 1108) show 55% aa identity in 170 aa overlap (BLASTp):

30	ORF3	3	IYLIRKNLGSVPVFFQERPGKDGPVKPMVKFRSMRDGLYSDGIPLPDGERLTPFGKKLRA	62
	yvfc	27	I + + R + GSPVFF Q RPG GKPF + KFR + M D' S G LPD RLT G + + R IAVVRLKIGSPVFFKQVRPGLHGKPF ¹ TYKFRTMTDERDSKGNLLPDEVRLTKTGR ² LIRK	86
	ORF3	63	ASXDELP ¹ ELWNILKGEMSLVGP ² RPLLMQYLPLYDNFQ ³ NR ⁴ RHEMKPGITGWAQVNGRNALS	122
	yvfc	87	S DELP + L N + LKG + + SLVGP ² RPLLM YLPLY Q RRHE + KPGITGWAQ + NGRNA + S LSIDELPQLLNVLKGDLSLVGP ² RPLLM DYLP ³ YTEQARRHEV ⁴ KPGITGWAQINGRNAIS	146
35	ORF3	123	WDEKFACDVWYIDHFS ¹ LC ² L ³ DX ⁴ XXXXXXXXXXXXXXXXXEGISAQGE ⁵ XTMP ⁶ PFTG	172
	yvfc	147	W + + KF DVWY + D + + S LD EGI T FTG WEKKFELDVWYVDNWS ¹ FFLD ² LKILCLTV ³ RKVLVSEGIQ ⁴ TN ⁵ HVTAER ⁶ FTG	196

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 (SEQ ID NO: 12) shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) (SEQ ID NO: 18) from *N. gonorrhoeae*:

	orf3	ILIIYLIRKNLGSPVFFFQERPGKDGPVKMKFR	34
	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKNKGSPVFFIRERPGKDGPVKMKFR	60
5	orf3	SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRLMLQYLPL	94
	orf3ng	SMRDALDSDGIPLPDSERLTDGKKLRATSLDELPELWNVLKGEMSLVGPRLMLQYLPL	120
	orf3	YDNFQNRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLLDIKILLTVKKVL	154
	orf3ng	YNKFQNRHEMKPGITGWAQVNGRNALSWDEKFSQDVWYTDNFSFWLDMKILFLTVKKVL	180
10	orf3	IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng	IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAEALAAALGTIGEIVFLDDRTQGSVNG	240
15	orf3	FSVIGTTLLENLSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	274
	orf3ng	FPVIGTTLLENLSLSPEQFDITVAVGNNRIRRQITENAAALGFKLPVLVHPDATVSPSAI	300
	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKDGIVNTAATVDHCLLDFAVHISPGAHLGSGNTRIGEESR	360

20 The complete length ORF3ng nucleotide sequence [<SEQ ID 17>] (SEQ ID NO: 17) is:

	1	ATGAGTAAAG	CCGTCAAACG	CCTGTTCGAC	ATCATCGCAT	CCGCATCGGG
	51	GCTGATTGTC	CTGTCGCCCC	TGTTTTTGGT	TTAATATAC	CTCATCCGCA
	101	AAACTTAGG	TTCGCCCCGTC	TTCTTCattc	GGGAACGCCc	cgGAAAGGAc
25	151	ggaaaacCTT	TTAAATGGT	CAAATTCGT	TCCatgcgcg	acgcgcttGA
	201	TTCAGACGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTTCGGCA
	251	AAAAATTACG	CGCCACCACT	TTGGACGAAC	TTCTGAATT	ATGGAATGTC
	301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
	351	TCTGCCGCTT	TACAACAAAT	TTCAAAACCG	CCGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
30	451	GAAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTCA	GCTTTTGGCT
	501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
	551	GCATTTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
	601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
35	701	CCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCCGCACGAC	GCTGCTGCTT
	751	GAAAACAGTT	TATCGCCCCG	ACAATTCGAC	ATCACCGTCG	CCGTCGGCAA
	801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
	851	AACTGCCCGT	TCTGATTCAT	CCCGACGCGA	CCGTCTCGCC	TTCTGCAATA
	901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCGTAC	AGGCCGGCAG
40	951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
	1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCGGGCGC	GCACCTGTCTG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
	1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgcgcgGT	GCAGGGgcgGG
	1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTGCGGGG	CAACCCGGCA
45	1201	AAGCCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence [<SEQ ID 18>] (SEQ ID NO: 18):

	1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKNLGSPV	FFIRERPGKD
	51	GKPFKMKVFR	SMRDALDSDG	IPLPDSERLT	DFGKKLRATS	LDELPELWNV
50	101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRHE	MKPGITGWAQ	VNGRNALSWD
	151	EKFSCDVWYT	DNFSFWLDMK	ILFLTVKKVL	IKEGISAQGE	ATMPPFAGNR

201 KLAIVIGAGGH GKVVAELAAA LGTYGEIVFL DDRTQGSVNG FPVIGTLLLL
 251 ENSLSPEQFD ITVAVGNNRI RRQITENAAA LGFKLPVLIH PDATVSPSAI
 301 IGQGSVVMAK AVVQAGSVLK DGVIVNTAAT VDHDCLLD AFVHISPGAHLS
 351 GNTRIGEESR IGTGACSRQQ TTVGSGVTAG AGAVIVCDIP DGMTVAGNPA
 401 KPLTGKNPKT GTA*

This protein shows 86.9% identity in 413 aa overlap with ORF3-1 (SEQ ID NO: 14):

10	orf3-1.pep	10 20 30 40 50 60	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKNLGSPVFFQERPGKDGKPFKMKVFR
	orf3ng	10 20 30 40 50 60	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLRKNLGSPVFFIRERPGKDGKPFKMKVFR
15	orf3-1.pep	70 80 90 100 110 120	SMRDALDSDGIPDPGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL
	orf3ng	70 80 90 100 110 120	SMRDALDSDGIPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL
20	orf3-1.pep	130 140 150 160 170 180	YDNFQNRHMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLLDIKILLTVKKVL
	orf3ng	130 140 150 160 170 180	YNKFQNRHMKPGITGWAQVNGRNALSWDEKFSQCDVWYTDNFSFWLDMKILFLTVMKKVL
25	orf3-1.pep	190 200 210 220 230 240	IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVDLAAALGRYREIVFLDDRAQGSVNG
	orf3ng	190 200 210 220 230 240	IKEGISAQGEATMPPTAGNRKLAVIGAGGHGKVVAELAAALGTYGEIVFLDDRTQGSVNG
30	orf3-1.pep	250 260 270 280 290 300	FSVIGTLLLENSLSPEQYDVA VAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT
	orf3ng	250 260 270 280 290 300	FPVIGTLLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFALPVLVHPDATVSPSAI
35	orf3-1.pep	310 320 330 340 350 360	VGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEESW
	orf3ng	310 320 330 340 350 360	IGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLD AFVHISPGAHLSGNTRIGEESR
40	orf3-1.pep	370 380 390 400 410	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX
	orf3ng	370 380 390 400 410	IGTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTX

In addition, ORF3ng (SEQ ID NO: 18) shows significant homology with a hypothetical protein

(SEQ ID NO: 1110) from *B.subtilis*:

gnl|PID|e238668 (Z71928) hypothetical protein [Bacillus subtilis]
)gi|1945702|gnl|PID|e313004 (Z94043) hypothetical protein [Bacillus subtilis]
)gi|2635938|gnl|PID|e1186113 (Z99121) similar to capsular polysaccharide
 biosynthesis [Bacillus subtilis] Length = 202

Score = 235 bits (594), Expect = 3e-61
Identities = 114/195 (58%), Positives = 142/195 (72%)

5 Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLIRKNLGSPPVFFIRERPGKDGPFPKMKVFRSMRD 64
+KRLFD+ A+ L S + L I ++R +GSPVFF + RPG GKPF + KFR+M D
Sbjct: 3 LKRLFDLTAAIFLLCCTSVIILFTIAVVRLKIGSPVFFKQVRPGLHGKPFPTLYKFRTMTD 62

Query: 65 ALDSGDIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPLYNKF 124
DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRPLLM YLPLY +
Sbjct: 63 ERDSKGNLLPDEVRLTKTGRLLIRKLSIDELPQLLNVLKGDLVLVGPRPLLMQYLPLYNKF 122

10 Query: 125 QNRRHEMKPGITGWAQVNGRNALSWDEKFSQVWYTDNFSFWLDMKILFLTVKKVLIKEG 184
Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG
Sbjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDLKLCTVRKVLVSEG 182

Query: 185 ISAQGEATMPFPFAGN 199
I T F G+
Sbjct: 183 IQQTNHVTAERFTGS 197

- 15 The hypothetical product of *yvf* gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 4

- 20 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 19>] (SEQ ID NO: 19):

25 1 ..AACCATATGG CGATTGTCAT CGACGAATAC GGC GG CACAT CCGGCTTGTT
51 CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT
101 TTGACGAAGA CGATAGCGCC GACAAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301 TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA
30 401 CAGTTTAG

This corresponds to amino acid sequence [<SEQ ID 20; ORF5>] (SEQ ID NO: 20; ORF5):

35 1 ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51 WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTSA RARRKSPYRR
101 FAVHRRTRRQ PPPAYADGDP REVS....XR RFCTV*

Further sequence analysis revealed the complete DNA sequence to be [<SEQ ID 21>] (SEQ ID NO: 21):

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85

1 ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCC
51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTTCGGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTTAAACCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA
451 CAGCGCAACC ATATGGCGAT TGTCATCGAC GAATACGGCG GCACATCCGG
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
651 CTTCCGCACG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC
701 ATTCAAGAGT TGGGACATCT GCCCGTGC GCAGAAAAG TCCTTATCGG
751 CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC
801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
851 TGACGGTACG GCGGTTTTCT GTTTCATCC GCCCATCCG CCAAACATAA

20 This corresponds to amino acid sequence [<SEQ ID 22; ORF5-1>] (SEQ ID NO: 22; ORF5-1):

25
30
35
40
45
50
55
60
65
70
75
80
85

1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLRLLE
51 KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEEADTI RPHGSRVGT ARARRKSPYR
251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*

Further work identified the corresponding gene in strain A of *N.meningitidis* [<SEQ ID 23>] (SEQ ID NO: 23):

30
35
40
45
50
55
60
65
70
75
80
85
90
95

1 ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCC
51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGTGAAGAC
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CTTGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA
451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG
501 TTTGGTAACT TTTGAAGACA TCATCGAGCA AATCGTCGGC GACATCGAAG
551 ATGAGTTTGA CGAAGACGAA AGCGCGGACA ACATCCACGC CGTTTCCGCC
601 GAACGCTGGC GCATCCACGC GGCTACCGAA ATCGAAGACA TCAACGCCTT
651 TTTCCGCACG GAATACAGCA GCGAAGAAGC CGACACCATC GGCGGCCNTG
701 GTCATTGAGG AATTGGNACA CCTGCCCGTG CGCGGCGAAA AAGTCNTTAT
751 CGGCGNNTTG CANTTCACNG TCGCCNGCGC NGACAACCGC CGCCTGCATA
801 CGCTGATGGC GACCCGCGTG AAGTAAGCTC CGCCGTTTCT GTACAGTTTA
851 GGATGACGGT ACGGGCGTTT TCTGTTTCAA TCCGCCCCAT CCGCCANACA
901 TAA

50 This encodes a protein having amino acid sequence [<SEQ ID 24; ORF5a>] (SEQ ID NO: 24; ORF5a):

1 MDGAQPKTNF XXRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLRLLE

51 KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
 101 KDEVLGILHA KDLLKYMFPN EQFHLKSILR PAVFVPEGKS LTALLKEFRE
 151 QRNHMAIVID EYGGTSGGLVT FEDIIEQIVG DIEDEFDEDE SADNIHAVSA
 201 ERWRIHAATE IEDINAFPGT EYSSEEADTI GGXGHSIGT PARARRKSY
 251 RRXAXHXRXR QPPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT
 301 *

The originally-identified partial strain B sequence (ORF5) (SEQ ID NO: 20) shows 54.7% identity over a 124aa overlap with ORF5a (SEQ ID NO: 24):

10 orf5.pep NHMAIVIDEYGGTSGGLVTFEDIIEQIVGEI
 orf5a FHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVGDI
 130 140 150 160 170 180

15 orf5.pep EDEFDEDDSDADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA
 orf5a EDEFDEDESADNIHAVSAERWRIHAATEIEDINAFPGTEYSSEEADTIGGXGHSIGTTPA
 190 200 210 220 230 240

20 orf5.pep RARRKSPYRRFAVHRRTTRQPPPAYADGDPREVSSXXXRRFCTV
 orf5a RARRKSXYRRXAXHXRXRQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXTX
 250 260 270 280 290 300

The complete strain B sequence (ORF5-1) (SEQ ID NO: 22) and ORF5a (SEQ ID NO: 24) show 92.7% identity in 300 aa overlap:

30 orf5a.pep MDGAQPKTNFXXRLIARLAREPDS AEDVLTLLRQAHEQEVFDADTLRLKVLDFSDLEV
 orf5-1 MDGAQPKTNFFERLIARLAREPDS AEDVLTLLRQAHEQEVFDADTLRLKVLDFSDLEV
 10 20 30 40 50 60

35 orf5a.pep RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN
 orf5-1 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN
 70 80 90 100 110 120

40 orf5a.pep EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG
 orf5-1 EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG
 130 140 150 160 170 180

45 orf5a.pep DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFPGTEYSSEEADTIGGXGHSIGT
 orf5-1 EIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT
 190 200 210 220 230
 250 260 270 280 290 300

```

orf5a.pep  PARARRKSXYRRXAXHXRXRQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXT
           ||||| ||| | | :| ||||| ||||| ||||| :|||: ||||| ||||| |||
orf5-1     SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT
           240      250      260      270      280      290

```

Further work identified the a partial DNA sequence in *N.gonorrhoeae* [SEQ ID 25] (SEQ ID NO: 25) which encodes a protein having amino acid sequence [SEQ ID NO: 26; ORF5ng] (SEQ ID NO: 26; ORF5ng):

```

1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
51 KVLDFAELEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFPN EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFPGT EYGSEEDTI RRLGHSIGT PARARRKSPY
251 RRFVHRRPR RQPPAHADG DPREVSRACP HRRFCTV*

```

Further analysis revealed the complete gonococcal nucleotide sequence [SEQ ID 27] (SEQ ID NO: 27) to be:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTTGAACGCC TGATTGCCCC
51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA ACAGGAAGTT TTTGATGCCG ACACACTGAC CCGGCTGGAA
151 AAAGTATTGG ACTTTGCCGA GCTGGAAGTG CGCGATGCGA TGATTACGCG
201 CAGCCGCATG AACGTATTGA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTCAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTGAAATC CGTCTTGCGC CCTGCCGTTT
401 TCGTGCCCGA AGGCAAATCT TTGACCGCCC TTTTAAAAGA GTTCCCGGAA
451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG
501 TTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGT GACATCGAAG
551 ACGAGTTTGA CGAAGACGAA AGCGccgacg acatCCACTC cgTTTccgCC
601 GAACGCTGGC GCATCCacgc ggctaCCGAA ATCGAAGaca TCAACGCCTT
651 TTTCCGTACG GAatacggca gcaagaagc cgacaccatc cggcggtTG
701 GTCATTCAGG AATTGGGACA CCTGCCCGTG CGCGGCGAAA AAGTCCTTat
751 cggcgGTTTg Cagttcaccg tCGCCCGCGC CGACAACCGC CGCCTGCACA
801 CGCTGATGGC GACCCGCGTG AAGTAAGCAG AGCCTGCCcg AccgcccgttT
851 CTGCacAGTT TAGGatgACG gtaCGGTCGT TTTCTGTTTC AATCCGCCCC
901 ATCCGCCAAA CATAA

```

This encodes a protein having amino acid sequence [SEQ ID 28; ORFng-1] (SEQ ID NO: 28; ORF5ng-1):

```

1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
51 KVLDFAELEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFPN EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFPGT EYGSEEDTI RRLGHSIGT PARARRKSPY
251 RRFVHRRPR RQPPAHADG DPREVSRACP TAVSAQFRMT VRSFVSIRP
301 IRQT*

```

The originally-identified partial strain B sequence (ORF5) (SEQ ID NO: 20) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng) (SEQ ID NO: 26):

-92-

orf5 NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI 30
 orf5ng FHLKSVLRPAV FVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI 182

orf5 EDEFDEDDSDADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA 90
 orf5ng EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSIGITPA 242

orf5 RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSX----RRFCTV 131
 orf5ng RARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSACPHRRFCTV 287

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) (SEQ ID NO: 22 & SEQ ID NO: 28) show 92.4% identity in 304 aa overlap:

orf5ng-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFAELEV
 orf5-1 MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFSDLEV

orf5ng-1.pep RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN
 orf5-1 RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN

orf5ng-1.pep EQFHLKSVLRPAV FVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG
 orf5-1 EQFHLKSVLRPAV FVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG

orf5ng-1.pep DIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSIGIT
 orf5-1 EIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT

orf5ng-1.pep PARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSACPTAVSAQFRMTVRSFVSIRP
 orf5-1 SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVS----TAVSAQFRMTVRAFSVSIRP

orf5ng-1.pep IRQTX
 orf5-1 IRQTX
 300

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

Homology with hemolysin homolog TlyC (accession U32716) (SEQ ID NO: 1111) of *H.influenzae*

ORF5 (SEQ ID NO: 20) and TlyC proteins (SEQ ID NO: 1111) show 58% aa identity in 77 aa overlap (BLASTp).

```

ORF5   2   HMAIVIDEYGGTSGLVTFEDIIEQIVGEIEDEFDEDDSDADNIHAVSSDTWRIHAATEIED 61
          HMAIV+DE+G  SGLVT EDI+EQIVG+IEDEFDE++ AD I  +S  T+  +  A  T+I+D
5  TlyC  166 HMAIVVDEFGAVSGLVTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD 224

ORF5   62   INTFFGTEYSIEEADTI 78
          N  F  T++  EE  DTI
TlyC  225 FNAQFNTDFDDEEVDTI 241

```

10 ORF5ng-1 (SEQ ID NO: 28) also shows significant homology with TlyC (SEQ ID NO: 1111):

```

SCORES      Init1:   301  Initn:   419  Opt:    668
Smith-Waterman score: 668;    45.9% identity in 242 aa overlap

      10      20      30      40      50
orf5ng-1.pep  MDGAQPKTNFFERLIARLAR-EPDSAEDVLNLLRQAHEQEVFDADTLTRLEK
15          |  ||: ||:| : | : |:::|:::|:::| : | : |
tlyc_haein  MNDEQQNSNQSENTKKPFFQSLFGRFFQGELKNREELVEVIRDSEQNDLIDQNTREMIEG
          10      20      30      40      50      60

      60      70      80      90     100     109
orf5ng-1.pep  VLDFAELEVRDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGE--DKDEVLGILH
20          |::|::|::| | | |::: : : |::|::|::|::| : |::|::|::|
tlyc_haein  VMEIAELRVRDIMIPRSQIIFIEDQQLNTCLNTIIESAHSRFPVIADADDRDNIVGILH
          70      80      90     100     110     120

      110     120     130     140     150     160
orf5ng-1.pep  AKDLLKYMf-NPEQFHLKSVLRPAVFPVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGL
25          |||||::: : | | |::|::|::|::| : : |::| : | ||||:|::|::|
tlyc_haein  AKDLLKFLREDAEVFDLSSLLRPVVIVPESKRVDRLKDFRSERFHMAIVVDEFGAVSGL
          130     140     150     160     170     180

      170     180     190     200     210     220
orf5ng-1.pep  VTFEDIIEQIVGDIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEAD
30          ||:|::|::|::|::|::|::| | | |::| : : : | |::|::| |::|::|::|
tlyc_haein  VTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQFNTDFDDEEVD
          190     200     210     220     230

      230     240     250     260     270     280
orf5ng-1.pep  TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACTAVSAQF
35          || | : : | | | :
tlyc_haein  TIGGLIMQTFGYLPKRGEIILKNLQFKVTSADSRRLIQLRVTPDEHLAEMNNVDEKSE
          240     250     260     270     280     290

```

Homology with a hypothetical secreted protein from *E.coli*:

40 ORF5a (SEQ ID NO: 24) shows homology to a hypothetical secreted protein (SEQ ID NO: 1112) from *E.coli*:

sp|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION
)gi|1778577 (U82598) similar to H. influenzae [Escherichia coli])gi|1786879
 (AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
 approx. 440 aa protein YTFL_HAEIN SW: P44717 [Escherichia coli] Length = 292

Score = 212 bits (533), Expect = 3e-54
 Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

Query: 2 DGAQPKTNFXXRLIARLAR-EPDSAEDVLTLLRQAHEQEVDADTLLRLEKVLDFSDLEV 60
 D K F L+++L EP + +++L L+R + + ++ D DT LE V+D +D V
 Sbjct: 10 DTISNKKGFFSLLLSQLFHHGEPKNRDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

Query: 61 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
 RD MI RS+M LK N +++ +I++AHSRFPVI EDKD + GIL AKDLL +M +
 Sbjct: 70 RDIMIPRSQMITLKRNTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIV 179
 E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
 Sbjct: 130 AEAFSMDKVLQRQAVVVPESKRVDRLKEFRSQRVHMAIVIDEFGGVSGLVTTIEDILELIV 189

Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEDT 229
 G+IEDE+DE++ D +S W + A IED N FGT +S EE DT
 Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVD 238

- 20 Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

- 25 ORF5-1 (SEQ ID NO: 22) (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 (SEQ ID NO: 22) is a surface-exposed protein, and that it is a useful immunogen.

30 Example 5

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 29>] (SEQ ID NO: 29):

1 ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTC
 51 GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC
 101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
 151 GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTCCCCG ACAGCTACGA
 201 AATCGATGCG GGCGGCAGTG ATTTGCAGAT TTACCAAACC GCCTACAAgG

5
251 GCGATGCAAC GCCGCCTGAA TGA^gGGCATG GGAAAGCAGG CAGGACGGGC
301 TGCCTTATAA AAACCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
401 CAACCGCCTG AAAATCGGTA TGCGCCTGCA AACCgAssCG TCCGTGATTT
451 ACGGCATGGG TGCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
501 CGCGACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC
551 GATTGCGCTG CCC..

This corresponds to the amino acid sequence [<SEQ ID 30; ORF7>] (SEQ ID NO: 30; ORF7):

10
1 MRGGRPD^SVT VQII^EGS^RFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
51 DAFSGNPE^GQ FFPDSYEIDA GGS^DLQIYQT AYKAMQRR^LN EAWESRQDGL
101 PYKNPYE^MLI MAXLVEKETG HEAXXD^HVAS V^FVNRLKIGM RLQ^TXXSVIY
151 G^MGAAYK^GKI RKADLR^RDTP YNTYTRGGLP PTPIALP..

15 Further sequence analysis revealed the complete DNA sequence [<SEQ ID 31>] (SEQ ID NO: 31):

20
1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCCGC
51 AGCCGTTTT^C GCCGCGCTGC TTTTGTTC TAAGGATAAC GGCAGGGCAT
101 ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCGTCCGT CGGCAGGAAA
151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTCAGCGGC
25
451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
501 CAGTGATTG CAGATTTACC AAACCGCCTA CAAGGCGATG CAACGCCGCC
551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
651 AGCCGACCGC GACCATGTCT CTTCCGTCTT CGTCAACCGC CTGAAAATCG
30
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGC GGCA
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCGATTGCG CTGCCCGGCA
851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT
901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATT
35
951 GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTTGAAA AAATAA

This corresponds to the amino acid sequence [<SEQ ID 32; ORF7-1>] (SEQ ID NO: 32; ORF7-1):

40
1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
51 LAEDRIVFSR HVLTAAYVL GVHNLHTGT YRLPSEVSAW DILQMRGGR
101 PDSVTVQIIE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
151 NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRR^LNEAWES RQDGLPYKNP
201 YEMLIMASLV EKETGHEADR DHVASV^FVNR LKIGMRLQTD PSVIYGMGAA
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301 FVSKMDGTGL SQFS^HDLTEH NAAVRKYILK K*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein encoded by yceg gene (accession P44270) (SEQ ID NO: 1113) of H.influenzae

ORF7 (SEQ ID NO: 30) and yceg proteins (SEQ ID NO: 1113) show 44% aa identity in 192 aa overlap:

5	ORF7	1	MRGGRPDSVTVQIIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG	55
	yceg	102	+ G+ V+ IEG F RK ++ P + K SNE++ A ++ + LNSGKEVQFNVKWIEGKTFKDWKDLLENAPHLVQTLKDKSNEEIFALLDLPDIGNLELK	161
10	ORF7	56	NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRRLNEAWESRQDGLPYKNPYEMLIMAXLV	115
	yceg	162	N EG +PD+Y +DL++ + + + M++ LN+AW R + LP NPYEMLI+A +V NVEGWLYPDTYNYTPKSTDLLELLKRSARMKKALNKAWNERDEDLPLANPYEMLILASIV	221
10	ORF7	116	EKETGHEAXXDHVASVFNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRRDTPYNTYT	175
	yceg	222	EKETG VASVF+NRLK M+LQT +VIYGMG Y G IRK DL TPYNTY EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV	281
10	ORF7	176	RGGLPPTPIALP	187
	yceg	282	GLPPTPIA+P IDGLPPTPIAMP	293

The complete length YCEG protein (SEQ ID NO: 1113) has sequence:

20

1	<u>MKKFLIAILL</u>	<u>LILILAGVAS</u>	<u>FSYYKMTEFV</u>	KTPVNVQADE	LLTIERGTTS
51	SKLATLFEQE	KLIADGKLLP	YLLKLKPELN	KIKAGTYSLE	NVKTVQDLLD
101	LLNSGKEVQF	NVKWIEGKTF	KDWRKDLENA	PHLVQTLKDK	SNEEFALLD
151	LPDIGQNLEL	KNVEGWLYPD	TYNYTPKSTD	LELLKRSAER	MKKALNKAWN
201	ERDEDLPLAN	PYEMLILASI	VEKETGIANE	RAKAVSFFIN	RLKAKMKLQT
251	DPTVIYGMGE	NYNGNIRKKD	LETKTPYNTY	VIDGLPPTPI	AMPESSSLQA
301	VANPEKTDYF	YFVADGSGGH	KFTRNLNEHN	KAVQEYLRWY	RSQKNAK

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF7 (SEQ ID NO: 30) shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) (SEQ ID NO: 34) from strain A of *N. meningitidis*:

[illegible]

-97-

```

orf7.pep      GMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALP
               |||||
orf7a         GMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVSKM
               250      260      270      280      290      300

5  orf7a       DGTGLSQFSHDLTEHNAAVRKYLKXX
               310      320      330

```

The complete length ORF7a nucleotide sequence [<SEQ ID 33>] (SEQ ID NO: 33) is:

```

10      1  ATGTTGAGAA  AATTGTTGAA  ATGGTCTGCC  GTTTTTTTGA  CCGTATCGGC
      51  AGCCGTTTTC  GCCGCGCTGC  TTTTCGTCCC  TAAAGACAAC  GGCAGGGCAT
     101  ACAGGATTAA  AATTGCCAAA  AACCAGGGTA  TTTCGTCCGT  CGGCAGGAAA
     151  CTTGCCGAAG  ACCGCATCGT  GTTCAGCAGG  CATGTTTTGA  CGGCGGCGGC
     201  CTACGTTTTG  GGTGTGCACA  ACAGCTGCA  TACGGGGACG  TACAGACTGC
     251  CTTCGGAAGT  GTCTGCTTGG  GATATCTTGC  AGAAAATGCG  CGGCGGCAGG
     301  CCGGATTCGG  TTACCGTGCA  GATTATCGAA  GGTCGCGT  TTTCGCATAT
     351  GAGGAAAGTC  ATCGACGCAA  CGCCCGACAT  CGAACACGAC  ACCAAAGGCT
     401  GGAGCAATGA  AAAACTGATG  GCGGAAGTTG  CCCCTGATGC  CTTCAGCGGC
     451  AATCCTGAAG  GGCAGTTTTT  CCCCACACGC  TACGAAATCG  ATGCGGGCGG
     501  CAGCGATTTA  CGGATTTACC  AAATCGCCTA  CAAGGCGATG  CAACGCCGAC
     551  TGAATGAGGC  ATGGGAAAGC  AGGCAGGACG  GGCTGCCTTA  TAAAAACCTT
     601  TATGAAATGC  TGATTATGGC  GAGCCTGATC  GAAAAGGAAA  CAGGGCATGA
     651  AGCCGACCGC  GACCATGTCG  CTTCCGTCTT  CGTCAACCGC  CTGAAAATCG
     701  GTATGCGCCT  GCAAACCGAC  CCGTCCGTGA  TTTACGGCAT  GGGTGCGGCA
     751  TACAAGGGCA  AAATCCGTAA  AGCCGACCTG  CGCCGCGACA  CGCCGTACAA
     801  CACCTACACG  CGCGGCGGTC  TGCCGCCAAC  CCCGATCGCG  CTGCCCAGCA
     851  AGGCGGCACT  CGATGCCGCC  GCCCATCCGT  CCGGTGAAAA  ATACCTGTAT
     901  TTCGTGTCCA  AAATGGACGG  TACGGGCTTG  AGCCAGTTCA  GCCATGATT
     951  GACCGAACAC  AACGCCGCCG  TTCGCAAATA  TATTTTGAAA  AAATAA

```

30 This is predicted to encode a protein having amino acid sequence [<SEQ ID 34>] (SEQ ID NO: 34):

```

35      1  MLRKLLKWSA  VFLTVSAAVF  AALLFVPKDN  GRAYRIKIAK  NQGISSVGRK
     51  LAEDRIVFSR  HVLTAAYVL  GVHNRLHTGT  YRLPSEVSAW  DILQKMRGGR
    101  PDSVTQIIE  GSRFSHMRKV  IDATPDIEHD  TKGWSNEKLM  AEVAPDAFSG
    151  NPEGQFFPDS  YEIDAGGSDL  RIYQIAYKAM  QRRLEAWES  RQDGLPYKNP
    201  YEMLIMASLI  EKETGHEADR  DHVASVFVNR  LKIGMRLQTD  PSVIYGMGAA
    251  YKGKIRKADL  RRDTPYNTYT  RGGLPPTPIA  LPGKAALDAA  AHPSGEKYLY
    301  FVSKMDGTGL  SQFSHDLTEH  NAAVRKYILK  K*

```

40 A leader peptide is underlined.

ORF7a (SEQ ID NO: 34) and ORF7-1 (SEQ ID NO: 32) show 98.8% identity in 331 aa overlap:

```

      10      20      30      40      50      60
orf7a.pep  MLRKLLKWSAVFLTVSAAVF AALLFVPKDN GRAYRIKIAKNQGISSVGRK LAEDRIVFSR
               |||||
45  orf7-1   MLRKLLKWSAVFLTVSAAVF AALLFVPKDN GRAYRIKIAKNQGISSVGRK LAEDRIVFSR
               10      20      30      40      50      60

      70      80      90     100     110     120
orf7a.pep  HVLTAAYVLGVHNRLHTGT YRLPSEVSAW DILQKMRGGR PDSVTQIIE GSRFSHMRKV
               |||||

```

5	orf7-1	HVLTAAYVLGVHNRHTGTYRLPSEVSAWDILQKMRGGRPDSTVTQIIEGSRFSHMRKV	70	80	90	100	110	120
	orf7a.pep		130	140	150	160	170	180
10	orf7-1	IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSEIDAGGSDLRIYQIAYKAM	130	140	150	160	170	180
	orf7-1		190	200	210	220	230	240
15	orf7a.pep	QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTD	190	200	210	220	230	240
	orf7-1		250	260	270	280	290	300
20	orf7a.pep	PSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLY	250	260	270	280	290	300
	orf7-1		310	320	330			
	orf7a.pep	FVSKMDGTGLSQFSHDLTEHNAAVRKYILKKX	310	320	330			
	orf7-1							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 (SEQ ID NO: 30) shows 94.7% identity over a 187aa overlap with a predicted ORF
 25 (ORF7.ng) (SEQ ID NO: 36) from *N. gonorrhoeae*:

30	orf7	MRGGRPDSTVTQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	60
	orf7ng		60
	orf7	FFPDSEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMAXLVEKETG	120
	orf7ng		120
35	orf7	HEAXDHVASVFNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGLP	180
	orf7ng		180
	orf7	PTPIALP	187
	orf7ng		236

An ORF7ng nucleotide sequence [<SEQ ID 35>] (SEQ ID NO: 35) is predicted to encode a protein
 40 having amino acid sequence [<SEQ ID 36>] (SEQ ID NO: 36):

```

1  MRGGRPDSTVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
51 DAFSGNPEGQ FFPDSEIDA GGSDLQIYQT AYKAMQRRLN EAWAGRQDGL
101 PYKNPYEMLI MASLIEKETG HEADRDHVAS VFNRLKIGM RLQTDPSVIY

```

151 GMGAAYKGKI RKADLRDTP YNTYTGGGLP PTRIALPGKA AMDAAAHPSG
201 EKLYLYFVSKM DGTGLSQFSH DLTEHNAAVR KYILKK*

Further sequence analysis revealed a partial DNA sequence of ORF7ng [<SEQ ID 37>] (SEQ ID

5 NO: 37):

```

1  ..taccgaatca AGATTGCCAA AAATCAGGGT ATTTTCGTCGG TCGGCAGGAA
51  ACTTGCCgaA GACCGCATCG TGTTTCAGCAG GCATGTTTTG ACAGCGGCGG
101 CCTACGTTTT GGGTGTGCAC AACAGGCTGC ATACGGGGAC gTACAGATTG
151 CCTTCGGAAG TGTCTGCTTG GGATATCTTG CAGAAAATGC GCGGCGGCAG
10  201 GCCGGATTCC GTTACCGTGC AGATTATCGA AGGTTCGCGT TTTTCGCATA
251 TGAGGAAAGT CATCGACGCA ACGCCCGACA TCGGACACGA CACCAAAGGC
301 TGGAGCAATG AAAAACTGAT GCGGGAAGTT GCGCCCGATG CCTTCAGCGG
351 CAATCCTGAA GGGCAGTTTT TTCCCAGACAG CTACGAAATC GATGCGGGCG
401 GCAGCGATTT GCAGATTTAC CAAACCGCCT ACAAGGCGAT GCAACGCCGC
15  451 CTGAACGAGG CATGGGCAGG CAGGCAGGAC GGGCTGCCTT ATAAAAACCC
501 TTATGAAATG CTGATTATGG CGAGCCTGAT CGAAAAGGAA ACGGGGCATG
551 AGGCCGACCG CGACCATGTC GCTTCCGTCT TCGTCAACCG CCTGAAAATC
601 GGTATGCGCC TGCAAACCGA CCCGTCCGTG ATTTACGGCA TGGGTGCGGC
651 ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGCGAC ACGCCGTACA
20  701 aCAccTAtac gggcgggggc ttgccgcaa cccggattgc gctgccccgc
751 Aagcgggcaa tggatgccgc cgcccaccg tccggcgAa aatacctgTa
801 tttcgtgtcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT
851 TGACCGAACA CAACGCCGCc gTcCGCAAT ATATTTTGAA AAAATAA

```

25 This corresponds to the amino acid sequence [<SEQ ID 38; ORFng-1>] (SEQ ID NO: 38;
ORF7ng-1):

```

1  ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
51  PSEVSAWDIL QKMRGGRPDS VTVQIEGSR FSHMRKVIDA TPDIGHDTKG
101 WSNEKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QTAYKAMQRR
30  151 LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFNRLKI
201 GMRLQTDPSV IYGMGAAYKG KIRKADLRD TPYNTYTGGG LPPTRIALPG
251 KAAMDAAHP SGEKLYLFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK*

```

ORF7ng-1 (SEQ ID NO: 38) and ORF7-1 (SEQ ID NO: 32) show 98.0% identity in 298 aa

35 overlap:

```

10      20      30      40      50      60
orf7-1.pep  KLLKWSAVFLTVSAAVF AALLFV PKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSRHVL
40  orf7ng-1                                YRIKIAKNQGISSVGRKLAEDRIVFSRHVL
10      20      30
70      80      90      100     110     120
orf7-1.pep  TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIEGSRFSHMRKVIDA
45  orf7ng-1  TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIEGSRFSHMRKVIDA
40      50      60      70      80      90
130     140     150     160     170     180
orf7-1.pep  TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR
orf7ng-1    TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR

```

-100-

		100	110	120	130	140	150
		190	200	210	220	230	240
5	orf7-1.pep	LNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFVNRLKIGMRLQTDPSV					
	orf7ng-1	LNEAWAGRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSV					
		160	170	180	190	200	210
		250	260	270	280	290	300
10	orf7-1.pep	IYGMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVS					
	orf7ng-1	IYGMGAAYKGKIRKADLRDTPYNTYTGGLPPTRIALPGKAAMDAAAHPSGEKYLYFVS					
		220	230	240	250	260	270
		310	320	330			
15	orf7-1.pep	KMDGTGLSQFSHDLTEHNAAVRKYILKKX					
	orf7ng-1	KMDGTGLSQFSHDLTEHNAAVRKYILKKX					
		280	290				

In addition, ORF7ng-1 (SEQ ID NO: 38) shows significant homology with a hypothetical *E.coli* protein (SEQ ID NO: 1114):

20	sp P28306 YCEG_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION gi 1787339 (AE000210) o340; 100% identical to fragment YCEG_ECOLI SW: P28306 but has 97 additional C-terminal residues [Escherichia coli] Length = 340 Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57 Identities = 20/87 (22%), Positives = 40/87 (45%)	
25	Query: 10 GISSVGRKLAEDRIVFSRHVLTAAYVLGVHNLHTGTYRLPSEVSAWDILQKMRGGRPD 69 G ++G +L D+I+ V + + GTYR +++ ++L+ + G+	Sbjct: 49 GRLALGEQLYADKIINRPRVFQWLLRIEPLSHFKAGTYRFTPQMTVREMLKLLESGKEA 108
	Query: 70 SVTVQIIEGSRFSHMRKVIDATPDIGH 96 ++++EG R S K + P I H	Sbjct: 109 QFPLRLVEGMRLSDYLKQLREAPYIKH 135
30	Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57 Identities = 84/155 (54%), Positives = 111/155 (71%)	
	Query: 120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLENEAWAGRQDGLPYKNPYEMLIMASLIEK 179 EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK	Sbjct: 158 EGWFWPDTWMYTANTTDVALLKRAHKKMVKAVDSAWEGRADGLPYKDKNQLVTMASIIEK 217
35	Query: 180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRDTPYNTYTGG 239 ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT	Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277
	Query: 240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274 GLPP IA PG ++ AAAHP+ YLYFV+ G	Sbjct: 278 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312
40		

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 6

- 5 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 39>] (SEQ ID NO: 39):

```

1  CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT
51  ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG
101 GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC
151 GAACGCGCAC GGCTTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC
201 GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGCAGGCG GGAACGGCTC
251 TGGCAACCTA TATGCTGATG TTGGAACGCA CAAAATCCCC CGAAGTCGCC
301 GAACGCGCCT TGGAAATGGC CGTGTGCTG AACCGGTTT AACAGGCGGA
351 AATGATTTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC
15  401 AAAACGGGCG GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG
451 CATCTGGACG GACGGGAAGA AGTGCTGGCT CAGGCGGACG AAGGACAG

```

This corresponds to the amino acid sequence [<SEQ ID 40; ORF9>] (SEQ ID NO: 40; ORF9):

```

20 1  ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFRRQ QRYSEEEIKN
51  ERARLAAVGE RVNQIFTLLG GETALQKGQA GTALATYMLM LERTKSPEVA
101 ERALEMAVSL NAFEQAEMIV QKWRQIEPIP GKAQKRAGWL RNVLRERGNQ
151 HLDGREEVLA QADEGQ

```

Further sequence analysis revealed the complete DNA sequence [<SEQ ID 41>] (SEQ ID NO: 41):

```

25 1  ATGTTACCTA ACCGTTTCAA AATGTAACT GTGTTGACGG CAACCTTGAT
51  TGCCCGACAG GTATCTGCCG CCGGAGGCGG TGCGGGGGAT ATGAAACAGC
101 CGAAGGAAGT CGGAAAGGTT TTCAGAAAGC AGCAGCGTTA CAGCGAGGAA
151 GAAATCAAAA ACGAACGCGC ACGGCTTGCG GCAGTGGGCG AGCGGGTTAA
30 201 TCAGATATTT ACGTTGCTGG GAGGGGAAAC CGCCTTGCAA AAGGGGACAG
251 CGGGAACGGC TCTGGCAACC TATATGCTGA TGTGGAACG CACAAAATCC
301 CCCGAAGTCG CCGAACGCGC CTTGGAAATG GCCGTGTCGC TGAACGCGTT
351 TGAACAGGCG GAAATGATTT ATCAGAAATG GCGGCAGATT GAGCCTATAC
401 CGGGAAGGCG GCAAAAACGG GCGGGGTGGC TGCGGAACGT GCTGAGGGAA
451 AGAGGAAATC AGCATCTGGA CCGACTGGAA GAAGTGCTGG CTCAGGCGGA
35 501 CGAAGGACAG AACCAGAGGG TGTTTTATT GTTGGCACA GCGGCCGTGC
551 AACAGACGGG GTTGGCGCAA AAAGCATCGA AAGCGGTTCC CCGCGCGGCG
601 TTGAAATATG AACATCTGCC CGAAGCGGCG GTTGCCGATG TGGTGTTTCA
651 CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTTG CAGCGTTTGG
701 CGAAGCTCGA TACGGAATA TTGCCCCCA CTTAATGAC GTTGCCTCTG
40 751 ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA
801 CACCCAAAAC CTTTCGGCCG TCTGGCAGGA AATGGAAATT ATGAATCTGG
851 TTTCCCTGCA CAGGCTGGAT GATGCCTATG CGCGTTTGAA CGTGCTGTG
901 GAACGCAATC CGAATCGAGA CCTGTATATT CAGGCAGCGA TATTGGCGGC
951 AAACCGAAAA GAAGTGCTT CCGTTATCGA CGGCTACGCC GAAAAGGCAT
45 1001 ACGGCAGGGG GACGGAGGAA CAGCGGAGCA GGGCGGCGCT AACGGCGGCG
1051 ATGATGTATG CCGACCGCAG GGATTACGCC AAAGTCAGGC AGTGGCTGAA
1101 AAAAGTATCC GCGCCGAAT ACCTGTTCGA CAAAGGTGTG CTGGCGGCTG

```

5
10
15

```

1151 CGGCGGCTGT CGAGTTGGAC GGCAGGAGGG CGGCTTTGCG GCAGATCGGG
1201 AGGGTGCGGA AACTTCCCGA ACAGCAGGGG CGGTATTTTA CGGCAGACAA
1251 TTTGTCCAAA ATACAGATGC TCGCCCTGTC GAAGCTGCCC GATAAACGGG
1301 AGGCTTTGAG GGGGTTGGAC AAGATTATCG AAAAACCGCC TGCCGGCAGT
1351 AATACAGAGT TACAGGCAGA GGCATTGGTA CAGCGGTCAG TTGTTTACGA
1401 TCGGCTTGGC AAGCGGAAAA AAATGATTTC AGATCTTGAA AGGGCGTTCA
1451 GGCTTGCACC CGATAACGCT CAGATTATGA ATAATCTGGG CTACAGCCTG
1501 CTGACCGATT CCAAACGTTT GGACGAAGGT TTCGCCCTGC TTCAGACGGC
1551 ATACCAAATC AACCCGGACG ATACCGCTGT CAACGACAGC ATAGGCTGGG
1601 CGTATTACCT GAAAGGCGAC GCGGAAAGCG CGCTGCCGTA TCTGCGGTAT
1651 TCGTTTGAAA ACGACCCCGA GCCCGAAGTT GCCGCCCAT TGGGCGAAGT
1701 GTTGTGGGCA TTGGGCGAAC GCGATCAGGC GGTTGACGTA TGGACGCAGG
1751 CGGCACACCT TACGGGAGAC AAGAAAATAT GCGGGGAAAC GCTCAAACGT
1801 CACGGCATCG CATTGCCCCA ACCTTCCCGA AAACCTCGGA AATAA

```

This corresponds to the amino acid sequence [<SEQ ID 42; ORF9-1>] (SEQ ID NO: 42; ORF9-1):

20
25
30

```

1  MLPNRFKMLT VLTATLIAGQ VSAAGGGAGD MKQPKEVGKV FRKQORYSEE
51  EIKNERARLA AVGERVNQIF TLLGGETALQ KGQAGTALAT YMLMLERTKS
101 PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGKAQKR AGWLRNVLRE
151 RGNQHLDEGL EVLAQADEGQ NRRVFLLLAQ AAVQDGLAQ KASKAVRRRA
201 LKYEHLPEAA VADVVFVQSG REKEKAIGAL QRLAKLDTEI LPPTLMTLRL
251 TARKYPEILD GFFEQTDTON LSAVWQEMEI MNLVSLHRLD DAYARLNVLL
301 ERNPNADLYI QAAILAANRK EGASVIDGYA EKAYGRGTEE QRSRAALTA
351 MMYADRRDYA KVRQWLKKVS APEYLFDKGV LAAAAVELD GGRAALRQIG
401 RVRKLPEQQG RYFTADNLSK IQMLALSKLP DKREALRGLD KII EKPPAGS
451 NTELQAEALV QRSVVYDRLG KRKKMISDLE RAFRLAPDNA QIMNNLGYSL
501 LTDSKRLDEG FALLQTAYQI NPDDTAVNDS IGWAYYLGKD AESALPYLRY
551 SFENDPEPEV AAHLGEVLWA LGERDQAVDV WTQAAHLTGD KKIWRETLKR
601 HGIALPQPSR KPRK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF9 (SEQ ID NO: 40) shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) (SEQ ID NO: 44) from strain A of *N. meningitidis*:

35
40
45

```

              10      20      30      40      50
orf9.pep      RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQORYSEEEIKNERARLA
              || :||:|:|:|:|:|:| || |::| | ||||| ||||| ||||| |||||
orf9a          MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEVGKVFRKQORYSEEEIKNERARLA
              10      20      30      40      50

              60      70      80      90     100     110
orf9.pep      AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
              ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf9a          AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
              60      70      80      90     100     110

              120     130     140     150     160
orf9.pep      EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ
              ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf9a          EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRVFLLLAQ
              120     130     140     150     160     170

```

orf9a AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVSQXREKEKAIGALQRLAKLDTEI
 180 190 200 210 220 230

The complete length ORF9a nucleotide sequence [<SEQ ID 43>] (SEQ ID NO: 43) is:

```

5      1  ATGTTACCCG CCCGTTTCAC CATTTTATCT GTGCTCGCGG CAGCCCTGCT
      51  TGC CGGGCAG GCGTATGCCG CCGGCGCGGC GGATGCGAAG CCGCCGAAGG
     101  AAGTCGGAAG GGT TTTT CAGA AAGCAGCAGC GTTACAGCGA GGAAGAAATC
     151  AAAAACGAAC GCGCAGCGCT TGC GGCAGTG GGCAGCGGG TTAATCAGAT
     201  ATTTACGTTG CTGGGANGGG AAACCGCCTT GCAAAAGGGG CAGGCGGGAA
10    251  CGGCTCTGGC AACCTATATG CTGATGTTGG AACGCACAAA ATCCCCCGAA
     301  GTCGCCGAAC GCGCCTTGGA AATGGCCGTG TCNCTGAACG CGTTTGAACA
     351  GGC GGAATG ATTTATCAGA AATGGCGGCA GATTGAGCCT ATACCGGGTA
     401  AGGCGCAAAA ACGGGCGGGG TGGCTGCGGA ACGTGCTGAG GGAAAGAGGA
     451  AATCAGCATC TAGACGGACT GGAAGAANTG CTGGCTCAGG CGGACGAANG
15    501  ACAGAACCGC AGGGTGTTTT TATTGTTGGC ACAAGCCGCC GTGCAACAGG
     551  ACGGGTTGGC GCAAAAAGCA TCGAAAGCGG TTCGCCGCGC GGCGTTGAGA
     601  TATGAACATC TGCCCGAAGC GCGGGTTGCC GATGTGGTGT TCAGCGTACA
     651  GGNACGCGAA AAGGAAAAGG CAATCGGAGC TTTGCAGCGT TTGGCGAAGC
     701  TCGATACGGA AATATTGCCC CCCACTTTAA TGACGTTGCG TCTGACTGCA
20    751  CGCAAATATC CCGAAATACT CGACGGCTTT TTCGAGCAGA CAGACACCCA
     801  AAACCTTTTC GCGTCTGGC AGGAAATGGA AATTATGAAT CTGGTTTCCC
     851  TGACACGGCT GGATGATGCC TATGCGCGTT TGAACGTGCT GTTGGAACGC
     901  AATCCGAATG CAGACCTGTA TATTCAGGCA GCGATATTGG CGGCAAACCG
     951  AAAAGAANGT GCTTCCGTTA TCGACGGCTA CGCCGAAAAG GCATACGGCA
25   1001  GGGGGACGGG GGAACAGCGG GGCAGGGCGG CAATGACGGC GGCATGATA
     1051  TATGCCGACC GAAGGGATTA CACCAAAGTC AGGCAGTGGT TGAAAAAAGT
     1101  GTCCGCGCCG GAATACCTGT TCGACAAAGG TGTGCTGGCG GCTGCGGCGG
     1151  CTGTCGAGTT GGACNGCGGC AGGGCGGCTT TGCGGCAGAT CGGCAGGGTG
30   1201  CGGAAACTTC CCGAACAGCA GGGGCGGTAT TTACGGCAG ACAATTTGTC
     1251  CAAAATACAG ATGTTTCGCC TGTCGAAGCT GCCCGACAAA CGGGAGGCTT
     1301  TGAGGGGGTT GGACAAGATT ATCGAAAAAC CGCCTGCCGG CAGTAATACA
     1351  GAGTTACAGG CAGAGGCATT GGTACAGCGG TCAGTTGTTT ACGATCGGCT
     1401  TGGCAAGCGG AAAAAAATGA TTTCAGATCT TGAAAGGGCG TTCAGGCTTG
     1451  CACCCGATAA CGCTCAGATT ATGAATAATC TGGGCTACAG CCTGCTTTCC
35   1501  GATTCCAAAC GTTTGGACGA AGGCTTCGCC CTGCTTCAGA CGGCATACCA
     1551  AATCAACCCG GACGATACCG CTGTCAACGA CAGCATAGGC TGGGCGTATT
     1601  ACCTGAAANG CGACGCGGAA AGCGCGCTGC CGTATCTGCG GTATTCTGTT
     1651  GAAAACGACC CCGAGCCCGA AGTTGCCGCC CATTTGGGCG AAGTGTTGTG
     1701  GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
40   1751  ACCTTACGGG AGACAAGAAA ATATGGCGGG AAACGCTCAA ACGTCACGGC
     1801  ATCGCATTGC CCCAACCTTC CCGAAAACCT CGGAAATAA

```

This encodes a protein having amino acid sequence [<SEQ ID 44>] (SEQ ID NO: 44):

```

45    1  MLPARFTILS VLAAALLAQ AYAAGAADAK PPKEVGKVFR KQORYSEEEI
      51  KNERARLAAV GERVNQIFTL LGXETALQKG QAGTALATYM LMLERTKSPE
     101  VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGKAQKRAG WLRNVLRERG
     151  NQHLDGLEEX LAQADEXQNR RVFLLLAQAA VQDGLAQKA SKAVRRAALR
     201  YEHLPEAAVA DVVFSVQXRE KEKAIGALQR LAKLDTEILP PTLMTLRLTA
50   251  RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLHRLDDA YARLNVLLER
     301  NPNADLYIQA AILAA NRKEX ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI
     351  YADRRDYTKV RQWLKKVSAP EYLFDKGVLA AAAAVELDXG RAALRQIGRV
     401  RKLPEQQGRY FTADNLSKIQ MFALSKLPDK REALRGLDKI IEKPPAGSNT
     451  ELQAEALVQR SVVYDRLGKR KKMISDLERA FRLAPDNAQI MNNLGYSLLS
55   501  DSKRLDEGFA LLQATYQINP DDTAVNDSIG WAYYLKXDAE SALPYLRYSF
     551  ENDPEPEVAA HLGFEVLWALG ERDQAVDVWT QAAHLTGDKK IWRETLKRHG
     601  IALPQPSRKP RK*

```

ORF9a (SEQ ID NO: 44) and ORF9-1 (SEQ ID NO: 42) show 95.3% identity in 614 aa overlap:

		10	20	30	40	50
	orf9a.pep	MLPARFTILSVLAAALLAGQAYAAAG--AADAKPPKEVGKVFQKQRYSEEEIKNERARLA				
5	orf9-1	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFQKQRYSEEEIKNERARLA				
		10	20	30	40	50
	orf9a.pep	60	70	80	90	100
	orf9a.pep	AVGERVNQIFTLTGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
10	orf9-1	AVGERVNQIFTLTGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
		70	80	90	100	110
	orf9a.pep	120	130	140	150	160
	orf9a.pep	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGLEEXLAQADEXQNRVFLLLAQ				
15	orf9-1	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGLEEVLAQADEGQNRVFLLLAQ				
		130	140	150	160	170
	orf9a.pep	180	190	200	210	220
	orf9a.pep	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVSQXREKEKAIGALQRLAKLDTEI				
20	orf9-1	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVVSQGREKEKAIGALQRLAKLDTEI				
		190	200	210	220	230
	orf9a.pep	240	250	260	270	280
	orf9a.pep	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL				
25	orf9-1	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL				
		250	260	270	280	290
	orf9a.pep	300	310	320	330	340
	orf9a.pep	ERNPNADLYIQAAIILANRKEKXASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYT				
30	orf9-1	ERNPNADLYIQAAIILANRKEGASVIDGYAEKAYGRGTGEQGRSRAALTAAMMYADRRDYA				
		310	320	330	340	350
	orf9a.pep	360	370	380	390	400
	orf9a.pep	KVRQWLKKVSAPEYLFDKGVLAAAAVELDXGRAALRQIGRVRKLPEQQGRYFTADNLSK				
35	orf9-1	KVRQWLKKVSAPEYLFDKGVLAAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK				
		370	380	390	400	410
	orf9a.pep	420	430	440	450	460
	orf9a.pep	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQRSVVYDRLGKRKKMISDLE				
40	orf9-1	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQRSVVYDRLGKRKKMISDLE				
		430	440	450	460	470
	orf9a.pep	480	490	500	510	520
	orf9a.pep	RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD				
45	orf9-1	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD				
		490	500	510	520	530
	orf9a.pep	540	550	560	570	580
	orf9a.pep	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR				

orf9-1	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR	
	550 560 570 580 590 600	
5	orf9a.pep	600 610 HGIALPQPSRKPRKX
	orf9-1	 HGIALPQPSRKPRKX
		610

Homology with a predicted ORF from *N.gonorrhoeae*

- 10 ORF9 (SEQ ID NO: 40) shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) (SEQ ID NO: 46) from *N. gonorrhoeae*:

		RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFQRKQRYSEEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVLPKEVGKVLKRKHRRYSEEEIKNERAR	58
15	orf9	LAAVGERVNIQIFTLGGETALQKQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
	orf9ng	LAAVGERVNRVFTLLGGETALQKQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
	orf9	QAEMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ	166
20	orf9ng	QAEMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPQSDYVHQPMIFLLL	178

The ORF9ng nucleotide sequence [<SEQ ID 45>] (SEQ ID NO: 45) was predicted to encode a protein having including acid sequence [<SEQ ID 46>] (SEQ ID NO: 46):

25	1	MIMLPARFTI	LSVLAAALLA	GQAYAAGAAD	VELPKEVGKV	LRKHRRYSEE
	51	EIKNERARLA	AVGERVNRVF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIQKWRQI	EPIPGEAQKP	AGWLRNVLKE
	151	GGNPHLDRLE	EVPAQSDYVH	QPMIFLLLQ	AAVQHGGVAQ	KPSKAVRPAA
	201	YNYEVLPETA	GADAVFCVQG	PQYEKAIQSF	PPCGRNPQTE	NIAPPFNELF
30	251	RPTARISP	LLQRFRTPEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

Further sequence analysis revealed the complete length ORF9ng DNA sequence [<SEQ ID 47>] (SEQ ID NO: 47):

35	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCCTCGCAG	CAGCCCTGCT
	51	TGCCCGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGAAA	GGTTTAAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
	201	GTTTACGCTG	TTGGGCGGTG	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
40	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA

5
10
15
20
25
30

```

351  GGC GGAAATG  ATTTATCAGA  AATGgcggca  gatcgagcct  ataCcggggtg
401  aggcgcacaaa accgGcgggG  tggctgcgga  acgtattgaa  ggaagggGGa
451  aaTCAGCATC  TGGAcgggtt  gaaagaggTG  CtggcgcaAT  cggacgatGT
501  GCAAAAAcgc  aggaTATTTT  TGCTGCTGGT  GCAAGCCGCC  GTGCagcagg
551  gTGGGGTGGC  TCAAAAAGCA  TCGAAAGCGG  TTCGcgtgc  GGcgttgaAG
601  TATGAACATC  TGCCcgaagc  ggcggTTGCC  GATGcggTGT  TCGGCGTACA
651  GGGACGCGAA  AAGGAAAagg  caaTCGAAGC  TTTGCAGCGT  TTGGCGAAGC
701  TCGATACGGA  AATATTGCCC  CCCACTTTAA  TGACGTTGCG  TCTGACTGCA
751  CGCAAATATC  CCGAAATACT  CGACGGCTTT  TTCGAGCAGA  CAGACACCCA
801  AAACCTTTCG  GCCGTCTGGC  AGGAAATGGA  AATTATGAAT  CTGGTTTCCC
851  TGCCTAAGCC  GGATGATGCC  TATGCGCGTT  TGAACGTGCT  GTTGAACAC
901  AACCCGAATG  CAAACCTGTA  TATTCAGGCG  GCGATATTGG  CGGCAAACCG
951  AAAAGAAGGT  GCGTCCGTTA  TCGACGGCTA  CGCCGAAAAG  GCATACGGCA
1001 GGGGGACGGG  GGAACAGCGG  GGCagggcgg  cAATgacggc  GGCGATGATA
1051 TATGCCGACC  GCAGGGATTA  CGCCAAAGTC  AGGCAGTGGT  TGAAAAAAGT
1101 GTCCGCGCCG  GAATACCTGT  TCGACAAAGG  CGTGCTGGCG  GCTGCGCGG
1151 CTGCCGAATT  GGACGAGGC  CGGGCGGCTT  TCGGCAGAT  CGGCAGGGTG
1201 CGGAAACTTC  CCGAACAGCA  GGGGCGGTAT  TTTACGGCAG  ACAATTTGTC
1251 CAAAATACAG  ATGCTCGCCC  TGTCGAAGCT  GCCCGACAAA  CGGGAAGCCC
1301 TGATCGGGCT  GAACAACATC  ATCGCCAAAC  TTTGCGCGGC  GGAAGCACG
1351 GAACCTTTGG  CGGAAGCATT  GGCACAGCGT  TCCATTATTT  ACGaacAGTT
1401 cggCAAACGG  GGA AAAATGA  TTGCCGACCT  tgaAAcgcg  CTCAAACCTTA
1451 CGCCCGATAA  TGCACAAATT  ATGAATAATC  TGGGCTACAG  CTTGCTTTCC
1501 GATTCCAAAC  GTTTTGACGA  GGGTTTCGCC  CTGCTTCAGA  CGGCATACCA
1551 AATCAACCCG  GACGATACCG  CCGTTAACGA  CAGCATAGGC  TGGGCGTATT
1601 ACCTGAAAGG  CGAGcgggaA  AGCGCGCTGC  CGTATCTGcg  gtattcgttt
1651 gAAAACGACC  CCGAGCCCGA  AGTTGCCGCC  CATTTGGGCG  AAGTGTGTGTG
1701 GGCATTGGGC  GAACGCGATC  AGGCGGTTGA  CGTATGGACG  CAGGCGGCAC
1751 ACCTTAGGGG  AGACAAGAAA  ATATGGCGGG  AGACGCTCAA  ACGCTACGGA
1801 ATCGCCTTGC  CCGAGCCTTC  CCGAAAACCC  CGGAAATAA

```

This encodes a protein having amino acid sequence [SEQ ID 48] (SEQ ID NO: 48):

35
40
45

```

1  MLPARFTILS  VLAAALLAGQ  AYAAGAADVE  LPKEVGKVLK  KHRRYSEEEI
51  KNERARLAAV  GERVNRVFTL  LGGETALQKG  QAGTALATYM  LMLERTKSPE
101 VAERALEMAV  SLNAFEQAEM  IYQKWRQIEP  IPGEAQKPAG  WLRNVLKEGG
151 NQHLDGLKEV  LAQSDDVQKR  RIFLLLVQAA  VQQGGVAQKA  SKAVRRAALK
201 YEHLPEAAVA  DAVFGVQGRE  KEKAIEALQR  LAKLDTEILP  PTLMTLRLTA
251 RKYPEILDGF  FEQTDTONLS  AVWQEMEIMN  LVSLRKPDDA  YARLNVLLEH
301 NPNANLYIQA  AILANRKEG  ASVIDGYAEK  AYGRGTGEQR  GRAAMTAAMI
351 YADRRDYAKV  RQWLKKVSAP  EYLFDKGVLA  AAAAAELDGG  RAALRQIGRV
401 RKLPEQQGRY  FTADNLSKIQ  MLALSCLPKD  REALIGLNNI  IAKLSAAGST
451 EPLAEALAQR  SIIYEQFGKR  GKMIADLETA  LKLTPDNAQI  MNNLGYSLLS
501 DSKRLDEGFA  LLQTA YQINP  DDTAVNDSIG  WAYYLGDAE  SALPYLRYSF
551 ENDPEPEVAA  HLGEVLWALG  ERDQAVDVWT  QAAHLRGDKK  IWRETLKRYG
601 IALPEPSRKP  RK*

```

ORF9ng (SEQ ID NO: 48) and ORF9-1 (SEQ ID NO: 42) show 88.1% identity in 614 aa overlap:

50
55

```

              10      20      30      40      50      60
orf9-1.pep  MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFVRKQQRYSEEEIKNERARLA
              |||  ||  :|:|:|:|:|:|  |||  |:|:|  |||||:|:|:|:|:|:|:|:|:|
orf9ng-1    MLPARFTILSVLAAALLAGQAYAAAG--AADVELPKEVGKVLKHHRRYSEEEIKNERARLA
              10      20      30      40      50

              70      80      90      100     110     120
orf9-1.pep  AVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
              |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
orf9ng-1    AVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA

```

		60	70	80	90	100	110
		130	140	150	160	170	180
5	orf9-1.pep	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRERGNQHLDGLEEVLAQADEGQNRVFLLLAQ					
	orf9ng-1	EMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNQHLDGLEEVLAQSDDVQKRRIFFLLVQ					
		120	130	140	150	160	170
		190	200	210	220	230	240
10	orf9-1.pep	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVVSQGREKEKAIGALQRLAKLDTEI					
	orf9ng-1	AAVQQGGVAQKASKAVRRAALKYEHLPEAAVADAVFGVQGREKEKAIEALQRLAKLDTEI					
		180	190	200	210	220	230
		250	260	270	280	290	300
15	orf9-1.pep	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVL					
	orf9ng-1	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKPDAYARLNVL					
		240	250	260	270	280	290
		310	320	330	340	350	360
20	orf9-1.pep	ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEEQSRRAALTAAMMYADRRDYA					
	orf9ng-1	EHNPNANLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQGRAAMTAAMYADRRDYA					
		300	310	320	330	340	350
		370	380	390	400	410	420
25	orf9-1.pep	KVRQWLKKVSAPEYLFDKGVLAATAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK					
	orf9ng-1	KVRQWLKKVSAPEYLFDKGVLAATAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK					
		360	370	380	390	400	410
		430	440	450	460	470	480
30	orf9-1.pep	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQRSVVYDRLGKRKKMISDLE					
	orf9ng-1	IQMLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQRSIIEYQFGKRKMIADLE					
		420	430	440	450	460	470
		490	500	510	520	530	540
35	orf9-1.pep	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD					
	orf9ng-1	TALKLTPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD					
		480	490	500	510	520	530
		550	560	570	580	590	600
40	orf9-1.pep	AESALPYLRYSFENDPEPEVAHLEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR					
	orf9ng-1	AESALPYLRYSFENDPEPEVAHLEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR					
		540	550	560	570	580	590
		610					
45	orf9-1.pep	HGIALPQPSRKPRKX					
	orf9ng-1	YGIALPEPSRKPRKX					
		600	610				

In addition, ORF9ng (SEQ ID NO: 48) shows significant homology with a hypothetical protein (SEQ ID NO: 1115) from *P.aeruginosa*:

sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION (ORF3)

)gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa)gi|557259 (X82071) orf3 [Pseudomonas aeruginosa] Length = 576

Score = 128 bits (318), Expect = 1e-28

Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

Query: 67 VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIQKWR 126
+++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
Sbjct: 53 LYSLLVAEELAGQRNRFDIASNYVVQAQKTRDPGVSEAFRIAIEYLGADQEALDTSLLWA 112

Query: 127 QIEPIPGEAQKPAG-----WLRNVLKEGGNQHLGLKEVLAQSDDVQKRRI 172
+ P +AQ+ A ++ VL G+ H D L A++D + +
Sbjct: 113 RSAPDNLDARAAAIQLARAGRYEESMVYMEKVLNGQGDTHFDLALSAETDPDTRAGL 172

Query: 173 FXXXXXXXXXXXXXXXXKASKAVRRAALKYEHLEAAVADAVFGVQGREKEKAIEALQRLA 232
++ KY + + A+ Q ++A+ L+ +
Sbjct: 173 L-----QSFHLLKKYPNNGQLLFGKALLQQDGRPDALTLEDNS 214

Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLRKP 287
E+ P L + L + K P + G E D + + + + LV +
Sbjct: 215 ASRHEVAPLLLSRLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL----LVEQNRL 270

Query: 288 DDAYARLNVLLLEHNPN-----ANLYIQAAI----- 312
DDA A L++ P+ A +Y++ +
Sbjct: 271 DDAKAEFAGLVQQFPDDDDDLRFSALVCLQAQWDEARIYLEELVERDSHVDAAHFNLG 330

Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371
LA +K+ A +D YA+ G G + T ++ A R D A R + P+
Sbjct: 331 RLAEQKDTARALDEYAAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388

Query: 372 YLFDKXXXXXXXXXXXXXXXXXQIGRVRLPEQQGRYFTADNLSKIQLALSKLPDKR 431
Y A L I+ ALS +
Sbjct: 389 Y-----AIQLYLIEAEALSNNDDQOE 408

Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
+A + + + E L L RS++ E+ +M DL + PDNA +
Sbjct: 409 KAWQAIQEGLKQYP-----EDL-NLLYTRSMLEAKRNDLAQMEKDLRFVIAREPDNAMAL 462

Query: 492 NNLGYSLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLLKGD AESALPYLRYSE 551
N LGY+L + R E L+ A++NPDD A+ DS+GW Y +G A YLR + +
Sbjct: 463 NALGYTLADRTTRYGEARELILKAHKLNPDDPAILD SMGWINYRQGLADAERYLRQALQ 522

Query: 552 NDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR 598
P+ EVA AHLGEVLWA G + A +W + + D + R T+KR
Sbjct: 523 RYPDHEVA AHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

gi|2983399 (AE000710) hypothetical protein (SEQ ID NO: 1116) [Aquifex aeolicus]
Length = 545

Score = 81.5 bits (198), Expect = 1e-14

Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

Query: 408 GRYFTADNL-SKIQMLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALQ----- 459
G Y A L K ++LA PDK+E L + +K + + L +
Sbjct: 335 GNYEDAKRLIEKAKVLA---PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390

Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513
+I+Y+ G L A++L P+N N LGYSLL +R++E L++

Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYLGYSLLLWYGKERVEEAEELIK 450

Query: 514 TAYQINPDDTAVNDSIGWAYYLKGD AESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572

A + +P++ A DS+GW YYLKG D E A+ YL + E +P V H+G+VL +G +

Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKG DYERAMQYLLKALREAYDDPVVNEHVG DVLLKMGYK 510

5 Query: 573 DQAVDVWTQAAHLRGDKK 590

++A + + +A L + K

Sbjct: 511 EEARNYYERALKLLEEGK 528

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 7

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 49>] (SEQ ID NO: 49):

```

1 AACCTCTACG CCGGCCCCGA GACCACATCC GTCATCGCAA ACATCGCCGA
51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
15 101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
151 TGGGCGATTA TCGTTTAAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGA CTGGGCG
20 351 GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
401 GCATTGTTCG CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTGGAT
451 TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCTGCCC ATCATTATGG
501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GACCGACCCG
551 ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTGTT
25 601 CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACACCTCC
651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
701 GCCCAAGGCG AAGTCGTTTC CTAA

```

This corresponds to the amino acid sequence [<SEQ ID 50; ORF11>] (SEQ ID NO: 50; ORF11):

```

30 1 ..NLYAGPQTTS VIANIADNLQ LAKDYGKVHW FASPLFWLLN QLHNIIGNWG
51 WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQQ
101 QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLGI
151 TDLSRADPYY ILPIIMAATM FAQTYLNPPP TDPMQAKMMK IMPLVFSXXF
35 201 FFFPAGXVLY WVNNLLTIA QQWHINRSIE KQRAQGEVVS *

```

Further sequence analysis revealed the complete DNA sequence [<SEQ ID 51>] (SEQ ID NO: 51):

```

1 ATGGATTTTA AAAGACTCAC GCGGTTTTTC GCCATCGCGC TGGTGATTAT
51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
40 101 AACAGGCAGC ACAACAACAG GCCGTAACCG CTTCCGCCGA AGCCGCGCTC
151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
201 TGATGAAAAA AGCGGCGACC TGCGCGGCTC GACCCTGCTC AAATACAAAG
251 CAACCGCGCA CGAAAATAAA CCGTTCATCC TGTTTGGCGA CGGCAAGAA
301 TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
351 TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGAAG

```

5
10
15
20
25

```

401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CGGTCTGAAA
451 ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
501 CTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACTG AGCGCGGACT
551 ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG TTACTIONTACC
601 CACTCTTACG TCGGCCCTGT TGTATTATACC CCTGAAGGCA ACTTCCAAAA
651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
701 CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC
751 CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAC AAAGCGTTTG
801 CGCCGCAGGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT
851 ACAGCACAGC CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA
901 GCCGAAGCCT CCATCAACCT CTACGCCGGC CCGCAGACCA CATCCGTCAT
951 CGCAAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAAGTAC
1001 ACTGGTTCGC CTCCCCTGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
1051 ATCGGCAACT GGGGTCTGGG GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAATGC
1151 GTGCCGCCGC ACCCAAACCTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACAAACAGG GATGATGCAG CTTTACACAG ACGAGAAAAT
1251 CAACCCGCTG GCGGCTGCTG TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401 GCCCATCATT ATGGCGGCAA CGATGTTTCG CCAAACCTAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCGTGGGT
1501 TTCTCCGTCA TGTCTCTCTT CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA
1601 TCGAAAAACA ACGCGCCCAA GCGGAAGTCG TTTCTTAA

```

This corresponds to the amino acid sequence [<SEQ ID 52; ORF11-1>] (SEQ ID NO: 52; ORF11-1):

30
35
40

```

1 MDFKRLTAFF AIALVIMIGW EKMFPPTPKPV PAPQQAQQQ AVTASAEAL
51 APATPITVTT DTVQAVIDEK SGDRLRLTLL KYKATGDENK PFILFGDGKE
101 YTYVAQSELL DAQGNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKVYFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVVYT PEGNFQKVSF SDLDDDAKSG KSEAERYIRKT PTGWLGMIEH
251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
301 AEASINLYAG PQTTSVIANI ADNLQLAKDY GKVHWFASPL FWLLNQLHNI
351 IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAPKL QAIKEKYGDD
401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451 WLGWITDLR ADPYYILPII MAATMFAQTY LNPPTDPMQ AKMMKIMPLV
501 FSVMFFFPPA GLVLYWVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) (SEQ ID NO: 1117) of *Pseudomonas putida*

45 ORF11 (SEQ ID NO: 50) and the 60kDa protein (SEQ ID NO: 1117) show 58% aa identity in 229 aa overlap (BLASTp).

```

ORF11 2 LYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61
      LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K
60K 324 LYAGPKIQSKLKELELTVLDYGFLEWFIQPIFWLLQHIHSLGNGWWSIIVLTMLIK 383

```

ORF11	62	AVLYPLTHASYSRSMAKMRAAAPKLQAIKEKYGDDRXXXXXXXXXXLYTDEKINPLGGCLPM	121
		+ +PL+ ASYSRMA+MRA APKL A+KE++GDDR LY EKinPLGGCLP+	
60K	384	GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKKINPLGGCLPI	443
ORF11	122	LLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQTYLNPPPT	181
		L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P	
60K	444	LVQMPVFLALYVWLLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQRLNPTTP	503
ORF11	182	DPMQAKMMKIMPLVXXXXXXXXXPAGXVLYVWVNNLLTIAQQWHINRSIE	230
		DPMQAK+MK+MP++ PAG VLYVWVNN L+I+QQW+I R IE	
60K	504	DPMQAKVMKMMPIIFTFFFLWFPAGLVLYVWVNNCLSSISQQWYITRRIE	552

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 (SEQ ID NO: 50) shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) (SEQ ID NO: 54) from strain A of *N. meningitidis*:

```

15      orf11.pep      10      20      30
                        NLYAGPQTTSVIANIADNLQLAKDYGKVHW
                        |||||
orf11a      IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQLXKDYGKVHW
                280      290      300      310      320      330

20      orf11.pep      40      50      60      70      80      90
                        FASPLFWLLNQLHNIIGNWGWAIIIVLTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKE
                        |||||
orf11a      FASPLFWLLNQLHNIIGNWGWAIIIVLTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKE
                340      350      360      370      380      390

25      orf11.pep      100      110      120      130      140      150
                        KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGIW
                        |||||
orf11a      KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGIW
                400      410      420      430      440      450

30      orf11.pep      160      170      180      190      200      210
                        TDLSRADPYYILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVFSXXXFFFPAGXVLY
                        |||||
orf11a      TDLSRADPYYILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVFSXXXFFFPAGLVLY
                460      470      480      490      500      510

35      orf11.pep      220      230      240
                        WVVNLLTIAQQWHINRSIEKQRAQGEVVSX
                        ||:|||||
orf11a      WVVNLLTIAQQWHINRSIEKQRAQGEVVSX
                520      530      540

```

The complete length ORF11a nucleotide sequence [[<SEQ ID 53>](#)] ([SEQ ID NO: 53](#)) is:

40	1	ANGGATTTTA	AAAGACTCAC	NGNGTTTTTC	GCCATCGCAC	TGGTGATTAT
	51	GATCGGATNG	NAAANGATGT	TCCCCACTCC	GAAGCCCGTC	CCCGCGCCCC
	101	AACAGACGGC	ACAACAACAG	GCCGTAANCG	CTTCGCGCCG	AGCCGCGCTC
	151	GCGCCCGNAN	CGCCGATTAC	CGTAACGCAC	GACACGGGTC	AAGCCGTCAT
	201	TGATGAAAAA	AGCGGCGCAC	TGCGCCGGCT	GACCCTGCTC	AAATACAAAG
45	251	CACCCGGCGA	CNAAAAATAA	CCGTTTCATC	TGTTTGGCGA	CGGCAAAANA

5
10
15
20
25

```

301 TACACCTACN TCGCCCATC CGAACCTTTG GACGCGCAGG GCAACAACAT
351 TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGAAG
401 GCGACAAAGT TGAAGTCCGC CTGAGCGCAC CTGAAACACG CGGTCTGAAA
451 ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
501 CTTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
551 ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG CTACTTTACC
601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
651 AGTCAGCTTC TCCGACTTGG ACGACGATGC CAANTCCGGN AAATCCGAGG
701 CCGAATACAT CCGCAAAACC CNGACCGGCT GGCTCGGCAT GATTGAACAC
751 CAGTTCATGT CCACCTGGAT CCTCCAACCC AAAGGCGGAC AAAGCGTTTG
801 CGCCGCTGGC GACTGCNGTA TNGACATCAA ACGCCGCAAC GACAAGCTGT
851 ACAGCACCAG CGTCAGCGTG CCTTAGCCG CTATCCAAAA CGGTGCGAAA
901 TCCNAAGCCT CCATCAACCT CTAGCCCGGC CCACAGACCA CATCNGTTAT
951 CGCAAACATC GCCGACAACC TGCAACTGNN CAAAGACTAC GGCAAAGTAC
1001 ACTGGTTCGC CTCCCCCTC TTTTGGCTTT TGAACCAACT GCACAACATC
1051 ATCGGCAACT GGGGTGCGG GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGTTCGATG GCGAAAATGC
1151 GTGCCGCCGC GCCCAAACCTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AGCAACAAGC CATGATGCAG CTTTACACAG ACGAGAAAAT
1251 CAACCGCTG GCGCGGTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCNT ACTACATCCT
1401 GCCCATCATT ATGGCGGCAA CGATGTTGCG CCAAACCTAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAATATCAT GCCTTTGGTT
1501 NTNTCNMNA NGTTCCTCNN CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551 GATCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCAGCA
1601 TCGAAAAACA ACGCGCCCAA GCGAAGTCG TTCCTAA

```

This encodes a protein having amino acid sequence [SEQ ID 54] (SEQ ID NO: 54):

30
35
40

```

1 XDFKRLTXFF AIALVIMIGX XXMFPTPKPV PAPQQTAAQQ AVXASAEAL
51 APXXPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDXNK PFILFGDGKX
101 YTYXAXSELL DAQGNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKVYFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVVYT PEGNFQKVSF SDLDDAXSG KSEAEYIRKT XTWLGMIEH
251 HFMSTWILQP KGGQSVCAAG DCXXDIKRRN DKLYSTSVSV PLAAIQNGAK
301 SXASINLYAG PQTTSVIANI ADNLQLXKDY GKVHWFASPL FWLLNQLHNI
351 IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAPKL QAIKEYGDD
401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451 WLGWITDLR ADPYYILPII MAATMFAQTY LNPPTDPMQ AKMMKIMPLV
501 XXXXFFXFP GLVLYWVINN LLTIAQQWHI NRSIEKQRAQ GEVVS*

```

ORF11a (SEQ ID NO: 54) and ORF11-1 (SEQ ID NO: 52) show 95.2% identity in 544 aa overlap:

45
50

```

              10      20      30      40      50      60
orf11a.pep    XDFKRLTXFFAIALVIMIGXXXMFPTPKVPVAPQQTAAQQAVXASAEALAPXXPITVTT
              |||||  |||||  |||||  |||||  |||||  |||||
orf11-1       MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVAPQQAQQAVTASAEALAPATPITVTT
              10      20      30      40      50      60

              70      80      90     100     110     120
orf11a.pep    DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFILFGDGKXYTYXAXSELLDAQGNILKG
              |||||  |||||  |||||  |||||  |||||  |||||
orf11-1       DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNILKG
              70      80      90     100     110     120

              130     140     150     160     170     180
orf11a.pep    IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYFTKGSYLVNVRFDIANGSGQTANL

```


-113-

	orf11-1	 IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL 130 140 150 160 170 180
5	orf11a.pep	190 200 210 220 230 240 SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAXSGKSEAHEYIRKT
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAHEYIRKT 190 200 210 220 230 240
10	orf11a.pep	250 260 270 280 290 300 XTGWLGMIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK 250 260 270 280 290 300
15	orf11a.pep	310 320 330 340 350 360 SXASINLYAGPQTTSVIANIADNLQLXKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV :
	orf11-1	AEASINLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV 310 320 330 340 350 360
20	orf11a.pep	370 380 390 400 410 420 LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL
	orf11-1	LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL 370 380 390 400 410 420
25	orf11a.pep	430 440 450 460 470 480 GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY 430 440 450 460 470 480
30	orf11a.pep	490 500 510 520 530 540 LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ 490 500 510 520 530 540
35	orf11a.pep	GEVVSX
	orf11-1	GEVVSX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 (SEQ ID NO: 50) shows 96.3% identity over a 240aa overlap with a predicted ORF

40 (ORF11.ng) (SEQ ID NO: 56) from *N. gonorrhoeae*:

Orf11	NLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	57
orf11ng	MAVNLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	60

	orf11	IIVKAVLYPLTNASYRSMAMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMAMRAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
5	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	180
	orf11	PPPTDPMQAKMMKIMPLVFSXXFFFPAGXVLYWVNNLLTIAQQWHINRSIEKQRAQGE	237
	orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGE	240
10	orf11	VVS 240	
	orf11ng	VVS 243	

An ORF11ng nucleotide sequence [<SEQ ID 55>] (SEQ ID NO: 55) was predicted to encode a
 15 protein having amino acid sequence [<SEQ ID 56>] (SEQ ID NO: 56):

	1	MAVNLYAGPQ	TTSVIANIAD	NLQLAKDYGK	VHWFASPLFW	LLNQLHNIIG
	51	NWGWAIVVLT	IIVKAVLYPL	TNASYRSMAM	MRAAAPELQT	IKEKYGDDRMA
	101	AQQQAMMQLF	EDEEINPLGG	CLPMLLQIPV	FIGLYWALFA	SVELRQAPWL
	151	GWITDLSRAD	PYYILPIIMA	ATMFAQTYLN	PPPTDPMQAK	MMKIMPLVFS
20	201	VMFFFPAGL	VLYWVNNLL	TIAQQWHINR	SIEKQRAQGE	VVS*

Further sequence analysis revealed the complete gonococcal DNA sequence [<SEQ ID 57>] (SEQ
 ID NO: 57) to be:

25	1	ATGGATTTTA	AAAGACTCAC	GGCGTTTTTC	GCCATCGCGC	TGGTGATTAT
	51	GATCGGCTGG	GAAAAAATGT	TCCCCACCCC	GAAACCCGTC	CCCGCGCCCC
	101	AACAGGCGGC	ACAAAAACAG	GCAGCAACCG	CTTCCGCCGA	AGCCGCGCTC
	151	GCGCCCGCAA	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTTAT
	201	TGATGAAAAA	AGTGGCGACC	TGCGCCGGCT	GACCTGCTC	AAATACAAAG
30	251	CAACCGCGCA	CGAAAACAAA	CCGTTCTGCC	TGTTTGGCGA	CGGCAAAGAA
	301	TACACCTACG	TCGCCCAATC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTGAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	ACCCTCAACG
	401	GCGACACAGT	CGAAGTCCGC	CTGAGCGCGC	CCGAAACCAA	CGGACTGAAA
	451	ATCGACAAAG	TCTATACCTT	TACCAAAGAC	AGCTATCTGG	TCAACGTCCG
35	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCgacTTgg	acgACGATGC	gaaaTccggc	aaATccgagg
	701	cgaatacaT	CCGCAAAACC	cgcaccggtt	ggctcggcat	gattgaacac
40	751	cacttcatgt	ccacctggat	cctccAAcct	aaaggcggcc	aaaaacgtttg
	801	cgccagggga	gactgccgta	tcgacattaa	aCgccgaac	gacaagctgt
	851	acagcgcaag	cgtcagcgtg	cctttaaccg	ctatcccaac	ccgggggcca
	901	aaaccgaaaa	tggcggTCAA	CCTGTATGCC	GGTCCGCAAA	CCACATCCGT
45	951	TATCGCAAAC	ATCGCcgacA	ACCTGCAACT	GGCAAAGAC	TACGGTAAAG
	1001	TACACTGGTT	CGCATCGCCG	CTCTTCTGGC	TCCTGAACCA	ACTGCACAAC
	1051	ATTATCGGCA	ACTGGGGCTG	GGCAATCGTC	GTTTGTACCA	TCATCGTCAA
	1101	AGCCGTACTG	TATCCATTGA	CCAACGcctc	ctACCGTTTCG	ATGGCGAAAA
	1151	TGCGTGccgc	cgcacCcaaA	CTGCAGACCA	TCAAAGAAAA	ATAcgGCGAC
	1201	GACCGTATGG	CGCAACAGCA	AGCGATGATG	CAGCTTTACA	AAGacgAGAA
50	1251	AATCAACCCG	CTGGGCGGCT	GTctgcctat	gctgttgCAA	ATCCCCGTCT
	1301	TCATCGGCTT	GTAATGGGCA	TTGTTTCGCT	CCGTAGAATT	GCGCCAGGCA
	1351	CCTTGGCTGG	GCTGGATTAC	CGACCTCAGC	CGCGCCGACC	CCTACTACAT

5
 1401 CCTGCCCATC ATTATGGCGG CAACGATGTT CGCCCAAACC TATCTGAACC
 1451 CGCCGCCGAC CGACCCGATG CAGGCGAAAA TGATGAAAAT CATGCCGTTG
 1501 GTTTTCTCCG TCATGTTCTT CTTCTTCCCT GCCGGTTTGG TTCTCTACTG
 1551 GGTGGTCAAC AACCTCCTGA CCATCGCCCA GCAGTGGCAC ATCAACCGCA
 1601 GCATCGAAAA ACAACGCGCC CAAGGCGAAG TCGTTTCCTA A

This encodes a protein having amino acid sequence [<SEQ ID 58; ORFng-1>] (SEQ ID NO: 58;
ORF11ng-1):

10
 1 MDFKRLTAF AIALVIMIGW EKMFPPTPKPV PAPQQAQKQ AATASAEAL
 51 APATPITVTT DTVQAVIDEK SGLRLRLTLL KYKATGDENK PFVLFPGDGKE
 101 YTYVAQSELL DAQGNILKG IGFSAPKKQY TLNGDTVEVR LSAPETNGLK
 151 IDKVYFTKD SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
 201 HSYVGPVVYT PEGNFQKVSF SDLDDAKSG KSEAEYIRKT PTGWLGMIEH
 251 HFMSTWILQP KGGQNVCAQG DCRIDIKRRN DKLYSASVSV PLTAIPTRGP
 15 301 KPKMAVNLYA GPQTTSVIAN IADNLQAKD YGKVHWFASP LFWLLNQLHN
 351 IIGNWGWAIV VLTIIIVKAVL YPLTNASYRS MAKMRAAAPK LQTIKEKYGD
 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA
 451 PWLGWITDLS RADPYIILPI IMAATMFAQT YLNPPTDPM QAKMMKIMPL
 501 VFSVMFFFFP AGLVLYWVWN NLLTIAQQWH INRSIEKQRA QGEVVS*

ORF11ng-1 (SEQ ID NO: 58) and ORF11-1 (SEQ ID NO: 52) shown 95.1% identity in 546 aa
 overlap:

25
 orf11ng-1.pep 10 20 30 40 50 60
 MDFKRLTAF AIALVIMIGW EKMFPPTPKPV PAPQQAQKQAATASAEALAPATPITVTT
 orf11-1 MDFKRLTAF AIALVIMIGW EKMFPPTPKPV PAPQQAQQA V TASAEALAPATPITVTT
 30
 orf11ng-1.pep 70 80 90 100 110 120
 DTVQAVIDEK SGLRLRLTLL KYKATGDENK PFVLFPGDGKEYTYVAQSELLDAQGNILKG
 orf11-1 DTVQAVIDEK SGLRLRLTLL KYKATGDENK PFILFGDGKEYTYVAQSELLDAQGNILKG
 35
 orf11ng-1.pep 130 140 150 160 170 180
 IGFSAPKKQY TLNGDTVEVRLSAPETNGLKIDKVYFTKDSYLVNVRFDIANGSGQTANL
 orf11-1 IGFSAPKKQY SLEGDKVEVRLSAPETRGLKIDKVYFTKGSYLVNVRFDIANGSGQTANL
 40
 orf11ng-1.pep 190 200 210 220 230 240
 SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT
 orf11-1 SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT
 45
 orf11ng-1.pep 250 260 270 280 290 300
 PTGWLGMIEHHFMSTWILQPKGGQNVCAQGD CRIDIKRRNDKLYSASVSVPLTAIPTRGP
 orf11-1 PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQN-GA
 orf11ng-1.pep 310 320 330 340 350 360
 KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIV

[illegible]

25 In addition, ORF11ng-1 (SEQ ID NO: 58) shows significant homology with an inner-membrane protein from the database (accession number p25754) (SEQ ID NO: 1117):

```

ID      60IM_PSEPU      STANDARD;      PRT;      560 AA.
AC      P25754;
DT      01-MAY-1992 (REL. 22, CREATED)
DT      01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT      01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE      60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES      Init1: 1074 Initn: 1293 Opt: 1103
Smith-Waterman score: 1406;      41.5% identity in 574 aa overlap

35
                                10          20          30          40
orf11ng-1.pep  MDFKR---LTAFFAIALVIMIGW-----EKMFEPT-----PKFPVPAPQQAQKQ
                ||:||      ::|| ::| |::| |      :  :||      |  ||| ::| :| :
p25754         MDIKRTILIAALAVVSYSVMVLKWNDDYGQAALPTQNTAASTVAPGLPDGVPAGNNGASAD
                10          20          30          40          50          60

40
                                50          60          70          80          90
orf11ng-1.pep  AATASAEAA LAPATPIT-----VTTDTVQAVIDEKSGDLRRLTLLKYKATGDE-NKPF
                :  :| :|| ::| |  :| ::      |  || ::| :||  :|| :| :| ||  | :| ||
p25754         VPSANAESSPAELAPVALSKDLIRVKTDVLELAIDPVGGDIVQLNLPKYPRRQDHPNIPF
                70          80          90          100          110          120

45
                                100          110          120          130          140
orf11ng-1.pep  VLFGDGKEYTYVAQSELLDAQGNILKGIG---FSAPKKQYTL-NGD---TVEVRLSAPE
                ||  :| | :| :||| |  :| :| :| :|  :| :| :| :| |  :| :| :| :| :
p25754         QLFDNGGERVYLAQSGLTGTGDPDA-RASGRPLYAAEQSKSYQLADGQEQLVVDLKFSS--

```

		130	140	150	160	170
	150	160	170	180	190	200
5	orf11ng-1.pep	TNGLKIDKVYTFTKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY				
	p25754	DNGVNYIKRFSFKRGEYDLNVSYLIDNQSGQAWNGNMFAQLKRDASGDPSSSTATGTATY				
		180	190	200	210	220
		230				
	210	220	230	240	250	260
10	orf11ng-1.pep	VGPVVYTPEGNFQKVSFSDLDLDDAKSGKSEAERYIKTPTGWLGMIEHHFMSTWILQPKGG				
	p25754	LGAALWTASEPYKKVSMKDID---KGSLE-----NVSGGWVAVLQHYFVTAWI-PAKSD				
		240	250	260	270	280
	270	280	290	300	310	320
15	orf11ng-1.pep	QNVCAQGDCRIDIKRRNDKLYSASVSPLTAIPTRGPKPKMAVNLYAGPQTTSVIANIAD				
	p25754	NNV-----VQTRKDSQGNIIIGYTGPIVSVPA-GGKVETSALLYAGPKIQSKLKELSP				
		290	300	310	320	330
	330	340	350	360	370	380
20	orf11ng-1.pep	NLQLAKDYGKVHWF-ASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYSMA				
	p25754	GLELTVDYGFL-WFIAQPIFWLLQHIHSLGNGWSIIVLTMLIKGLFFPLSAASYSMA				
		340	350	360	370	380
		390	400	410	420	430
25	orf11ng-1.pep	KMRAAAPKLQTIKEKYGDDRMAQQQAMQLYKDEKINPLGGCLPMLLQIPVFIGLYWALF				
	p25754	RMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPILVQMPVFLALYWVLL				
		400	410	420	430	440
		450				
	450	460	470	480	490	500
30	orf11ng-1.pep	ASVELRQAPWLGWITDLRADPPYIILPIIIMAAATMFAQTYLNPPPTDPMQAKMMKIMPLVF				
	p25754	ESVEMRQAPWILWITDLSEKDPFFILPIIMGATMFIQRLNPTPPDPMQAKVMKMMPIIF				
		460	470	480	490	500
		510	520	530	540	
35	orf11ng-1.pep	SVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGEVVSX				
	p25754	TFFFLWFPAGLVLYWVNNCLSSISQQWYITRRIEAAATKAAA				
		520	530	540	550	560

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 8

The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 59](#)] ([SEQ ID NO: 59](#)):

```

1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTNG
151 TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251 ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

10 This corresponds to the amino acid sequence [<SEQ ID 60; ORF13>] (SEQ ID NO: 60; ORF13):

```

1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVXY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

15 Further sequence analysis elaborated the DNA sequence slightly [<SEQ ID 61>] (SEQ ID NO: 61):

```

1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTnG
151 TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
251 ACCGTTACGA AGTTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

25 This corresponds to the amino acid sequence [<SEQ ID 62; ORF13-1>] (SEQ ID NO: 62; ORF13-1):

```

1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVFY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

30

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 (SEQ ID NO: 60) shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) (SEQ ID NO: 64) from strain A of *N. meningitidis*:

```

35
      10      20      30      40      50
orf13.pep  AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
      |||||||
orf13a     MTVWFVA AVAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
      10      20      30      40      50      60

      60      70      80      90     100     110
orf13.pep  VHAKTAVRKVETDSYQDLGAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA
      |||||||
orf13a     VHAKTAVGK VETDSYQDLGAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
      70      80      90     100     110     120

```

40

```

              120
orf13.pep    LIVRKEGNLLIITHPX
              |||||
orf13a       LIVRKEGNLLIIAKPX
              130

```

The complete length ORF13a nucleotide sequence [<SEQ ID 63>] (SEQ ID NO: 63) is:

```

1  ATGACTGTAT GGTGTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
51 GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
101 GCATTGCTTA CGGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
151 GCTCTGCTTT CCGCGCTGGG TATTTGGTTC GTACACGCCA AAACCGCCGT
201 GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
251 CCGAAATCCT CCGGCACGCA GCGGCAACC GTTACGAAGT TTTTATCGC
301 GGTACGCACT GGCAGGCTCA AAATACGGGG CAAGAAGAGC TTGAACCAGG
15 351 AACGCGCGCC CTAATCGTCC GCAAGGAAGG CAACCTTCTT ATCATCGCAA
401 AACCTTAA

```

This encodes a protein having amino acid sequence [<SEQ ID 64>] (SEQ ID NO: 64):

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAAAL AGSGIAYGLT GSTPAAVLTA
20 51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDA QYAEILRHA GGNRYEVFYP
101 GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP*

```

ORF13a (SEQ ID NO: 64) and ORF13-1 (SEQ ID NO: 62) show 94.4% identity in 126 aa overlap

```

25      10      20      30      40      50      60
orf13a.pep  MTVWFVAAVAVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
          |||||
orf13-1     AVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
          10      20      30      40      50

30      70      80      90      100     110     120
orf13a.pep  VHAKTAVGKVETDSYQDLDAQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
          |||||
orf13-1     VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
          60      70      80      90      100     110

35      130
orf13a.pep  LIVRKEGNLLIIAKPX
          |||||
orf13-1     LIVRKEGNLLIITHPX
          120

```

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF13 (SEQ ID NO: 60) shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) (SEQ ID NO: 66) from *N. gonorrhoeae*:

```

orf13          AVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF  51
          |||||
orf13ng        MTVWFVAAVAVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF  60

```

```

orf13      VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA 111
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf13ng    VHAKTAVGK VETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA 120

orf13      LIVRKEGNLLIITHP 126
          ||||| ||||| ||||| |||||
orf13ng    LIVRKEGNLLIIANP 135

```

The complete length ORF13ng nucleotide sequence [<SEQ ID 65>] (SEQ ID NO: 65) is:

```

1  ATGACTGTAT GGTTTGTGTC CGCTGTTGCC GTCTTAATCA TCGAATTATT
51 GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
101 GCATTGCCTA CGGGCTGACT GGCAGCACGC CTGCCGCCGT CTTGACCGCC
151 GCACTGCTTT CCGCGCTGGG CATTTGGTTC GTACATGCCA AAACCGCCGT
201 GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATACC GGAAAATATG
251 CCGAAATCCT CCGATACACA GCGGCAACC GTTACGAAGT TTTTATCGC
301 GGTACGCACT GGCAGGCGCA AAATACGGGG CAGGAAGTGT TTGAACCGGG
351 AACGCGGCC CTCATCGTCC GCAAAGAAGG TAACCTTCTT ATCATCGCAA
401 ACCCTTAA

```

This encodes a protein having amino acid sequence [<SEQ ID 66>] (SEQ ID NO: 66):

```

1  MTVWFVAAVA VLIIELLTGT VYLLVSAAL AGSGIAYGLT GSTPAAVLTA
51 ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYAEILRYT GGNRYEVFYR
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

ORF13ng (SEQ ID NO: 66) shows 91.3% identity in 126 aa overlap with ORF13-1 (SEQ ID NO: 62):

```

          10      20      30      40      50
orf13-1.pep  AVLIIELLTGT VYLLVSAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf13ng      MTVWFVAAVAVLIIELLTGT VYLLVSAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
          10      20      30      40      50      60

          60      70      80      90      100     110
orf13-1.pep  VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf13ng      VHAKTAVGK VETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA
          70      80      90      100     110     120

          120
orf13-1.pep  LIVRKEGNLLIITHPX
          ||||| ||||| ||||| |||||
orf13ng      LIVRKEGNLLIIANPX
          130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 (SEQ ID NO: 60) and ORF13ng (SEQ ID NO: 66) are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* [<SEQ ID 67>] (SEQ ID NO: 67):

```

1  ATGTWTGATT TCGGTTTrGG CGArCTGGTT TTTGTCGGCA TTATCGCCCT
5  51  GATwGtCCTC GGCCCCGAAC GCsTGCCGA GGCCGCCCGC AyCGCCGGAC
101 GGcTCATCGG CAGGCTGCAA CGCTTTGTCTG GcAGCGTCAA ACAGGAATTT
151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201 AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCCAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
10 351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
401 ATGCCGTC..

```

This corresponds to the amino acid sequence [<SEQ ID 68; ORF2>] (SEQ ID NO: 68; ORF2):

```

1  MXDFGLGELV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
15 51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGNPXS RCGKHPIRRH FRRYAV..

```

Further work revealed the complete nucleotide sequence [<SEQ ID 69>] (SEQ ID NO: 69):

```

20 1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
51  GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201 AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
25 251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCCAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
401 TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG
30 551 AAGTCAGCTA TATCGATACT GCTGTGAAA CGCCTGTTCC GCACACCACT
601 TCCCTGCGCA AACAGGCAAT AAGCCGAAA CGCGATTTC GTCCGAAACA
651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

```

This corresponds to the amino acid sequence [<SEQ ID 70; ORF2-1>] (SEQ ID NO: 70; ORF2-1):

```

35 1  MFDfGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGPLP DAANTLSDGI SDVMPSESY ASAETLGDSG
151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT
40 201 SLRKQAISRK RDRPKHRAK PKLRVRKS*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* [<SEQ ID 71>] (SEQ ID NO: 71):

```

45 1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
51  GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
151 GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA
201 AGCTGCCGCT GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA

```

-122-

251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCC GAAC AGCGCACGCC TGCTGATTTC GGTGTCGATG AAAACGGCAA
 351 TCCCTTTCCC GATGCGGCAA ACACCTATT AGACGGCATT TCCGACGTTA
 401 TGCCGTCCGA ACCTTCCTAC GCTTCCGCCG AAACCCCTGG GGACAGCGGG
 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG
 501 GCGGGAATAC CTGACTGCTT CTGCGCCGC ACCCGTCGTA CAGACCGTCG
 551 AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTCC GCATACCACT
 601 TCGCTGCGTA AACAGGCAAT AAGCCGCAA CGCGATTTCG GTCCTAAATC
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This encodes a protein having amino acid sequence [SEQ ID 72; ORF2a] (SEQ ID NO: 72; ORF2a):

1 MFD FGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNFP DAANTLLDGI SDVMPSESY ASAETLGDSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHHT
 201 SLRKQAIRSK RDLRPKSRAK PKLRVRKS*

The originally-identified partial strain B sequence (ORF2) (SEQ ID NO: 68) shows 97.5% identity over a 118aa overlap with ORF2a (SEQ ID NO: 72):

	10	20	30	40	50	60
orf2.pep	MXDFGLGELVFVGIIALIVL GPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR					
orf2a	MFD FGLGELVFVGIIALIVL GPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR					
	10	20	30	40	50	60
orf2.pep	KAKQFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEKLPEQRTPADFGVDENGPNXS					
orf2a	KAKQFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEKLPEQRTPADFGVDENGPNFP					
	70	80	90	100	110	120
orf2.pep	RCGKHPIRRHFRRYAV					
orf2a	DAANTLLDGISDVMPSESYASAETLGDSGQTGSTAEP AETDQDRAWREYLTASAAAPVV					
	130	140	150	160	170	180

The complete strain B sequence (ORF2-1) (SEQ ID NO: 70) and ORF2a (SEQ ID NO: 72) show 98.2% identity in 228 aa overlap:

orf2a.pep	MFD FGLGELVFVGIIALIVL GPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
orf2-1	MFD FGLGELVFVGIIALIVL GPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
orf2a.pep	KAKQFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEKLPEQRTPADFGVDENGPNFP	120
orf2-1	KAKQFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEKLPEQRTPADFGVDENGPNLP	120
orf2a.pep	DAANTLLDGISDVMPSESYASAETLGDSGQTGSTAEP AETDQDRAWREYLTASAAAPVV	180
orf2-1	DAANTLSDGISDVMPSESYASAETLGDSGQTGSTAEP AETDQDRAWREYLTASAAAPVV	180

```

orf2a.pep      QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAPKLRVRKSX 229
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf2-1         QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDFRPKHKRAKPKLRVRKSX 229

```

Further work identified a partial DNA sequence [<SEQ ID 73>] (SEQ ID NO: 73) in
5 *N.gonorrhoeae* encoding the following amino acid sequence [<SEQ ID 74; ORF2ng>] (SEQ ID
NO: 74; ORF2ng):

```

1  MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK
101 LPEQRTPADF GVDEKGNLSL RYGVKHIRRH FRRYAV*

```

Further work identified the complete gonococcal gene sequence [<SEQ ID 75>] (SEQ ID NO:
75):

```

1  ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTCCGCA TTATCGCCCT
51 GATTGTCCTT GGTCCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC
15 101 GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT
151 GACACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCGA
201 AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GATACGGATA
251 TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCCAAC AGCGCACGCG tgcgcgattc gGTGTCGATg AAAacggcaa
20 351 tcccccttccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA
401 TGCCGTCTGA ACGTTCCGAT ACTtccgcCG AAACCCTTGG GGACGACAGG
451 CAAACCGGCA GTACAGCGGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG
501 CGCGGAATAC CTGactgctt ctgcccgcgc acctgtcgta Cagagggccg
551 tcgaagtcag ctaTATCGAT ACTGCTGTTG AAacgcctgT tccgcaCacc
25 601 acttccctgc gcaAACAGGC AATAAACCGC AAACGCGATT TttgtccgaA
651 ACACCGCGCc aAACCGAAat tgcgcgtcCG TAAATCATAA

```

This encodes a protein having the amino acid sequence [<SEQ ID 76; ORF2ng-1>] (SEQ ID NO:
76; ORF2ng-1):

```

1  MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGPNLP DTANTVSDGI SDVMPSESD TSAETLGDDR
151 QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETPVPHT
35 201 TSLRKQAINR KRDFCPKHRA KPCLRVRKS*

```

The originally-identified partial strain B sequence (ORF2) (SEQ ID NO: 68) shows 87.5% identity
over a 136aa overlap with ORF2ng (SEQ ID NO: 74):

```

orf2.pep      MXDFGLGELVFVGIIALIVLGPXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
                |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 orf2ng      MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60

orf2.pep      KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNXS 120
                |:|:| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
orf2ng      KVKQAFEAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNLSL 120

```

```

orf2.pep      RCGKHPIRRHFRRYAV 136
              | | | | | | | | | |
orf2ng        RYGKHIRRHFRYAV 136

```

- 5 The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) (SEQ ID NO: 70 & SEQ ID NO: 76) show 91.7% identity in 229 aa overlap:

```

10 orf2-1.pep      10      20      30      40      50      60
    MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1          10      20      30      40      50      60
    MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR

15 orf2-1.pep      70      80      90      100     110     120
    KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNNPLP
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1          70      80      90      100     110     120
    KVKQAFEAAAQVRDSLKETDTMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNNPLP

20 orf2-1.pep      130     140     150     160     170     180
    DAANTLSDGISDVMPSPERSYASAETLGDSGQTGSTAEPATDQDRAWREYLTASAAAPVV
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1          130     140     150     160     170     180
    DTANTVSDGISDVMPSPERSDTSAETLGDDRQTGSTAEPATDKDRAWREYLTASAAAPVV

25 orf2-1.pep      190     200     210     220     229
    Q-TVEVSYIDTAVETVPVPHTTSLRKQAISRKRDFRPHKRAKPKLRVRKXS
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1          190     200     210     220     230
    QRAVEVSYIDTAVETVPVPHTTSLRKQAINRKRDPCPKHRAKPKLRVRKXS

```

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined), and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein
 30 (SEQ ID NO: 1118) of *E.coli*:

```

gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
Score = 56.6 bits (134), Expect = 1e-07
Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

```

```

35 Query: 1 MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
    MFD G EL+ V II L+VLGP+RLP A +T I L+ +V+ EL +++L+E +
Sbjct: 1 MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60

Query: 61 -KVKQAFEAAAQVRDSLKETDTMQNS 87
    +K+ +A+ + LK + +++ +
Sbjct: 61 DSLKKVEKASLTNLTPELKASMDLRQA 88
40

```

Based on this analysis, it was predicted that ORF2 (SEQ ID NO: 68), ORF2a (SEQ ID NO: 72) and ORF2ng (SEQ ID NO: 74) are likely to be membrane proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (SEQ ID NO: 70) (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis (Figure 3D). These experiments confirm that ORF37-1 (SEQ ID NO: 4) is a surface-exposed protein, and that it is a useful immunogen.

Example 10

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 77>] (SEQ ID NO: 77):

```

1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
51  CGC.TGCGGG AACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT
101 TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGG TCGCTACTCC
251 ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
351 GTTTGACAGG TTTAACCCTT TCTTTATCTA CACTTAATGC CCCTGCACTC
401 TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC
501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTCCTGCGC
551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

```

This corresponds to the amino acid sequence [<SEQ ID 78; ORF15>] (SEQ ID NO: 78; ORF15):

```

1  MQARLLIPII FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK
51  DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT
101 DYTYPRIYET AETTSGLTGT LTTSLSTLNA PALSRQSDG SGSKSSLGLN
151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201 IDVFGTIRNR TEM..

```

Further work revealed the complete nucleotide sequence [<SEQ ID 79>] (SEQ ID NO: 79):

```

1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
51  CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT
101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGTACTCCA
251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG

```

551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTT GCAGTAGACA
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
 801 AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTCTCC GATATCCGAC
 851 CATACGGCAA TCATACGGGT AACTCCGCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCGAC AACATAGACA
 951 AGGACAACCT TGA

This corresponds to the amino acid sequence [<SEQ ID 80; ORF15-1>] (SEQ ID NO: 80; ORF15-1):

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPREYET AETTSGLTGT LTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSIGIK PTEGLMVDFS DIRPYGNHTG NSAPSVEADN
 301 SHEGYGYSDE VVRQHRQGP *

Further work identified the corresponding gene in strain A of *N.meningitidis* [<SEQ ID 81>] (SEQ ID NO: 81):

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG ACGTACGAG GTATTCCATC GCATGGCGGA GGTAAACGCT
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
 551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTT GCAGTAGACA
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
 801 AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTCTCC GATATCCAAC
 851 CATACGGCAA TCATATGGGT AACTCTGCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA
 951 AGGGCAACCT TGA

This encodes a protein having amino acid sequence [<SEQ ID 82; ORF15a>] (SEQ ID NO: 82; ORF15a):

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPREYET AETTSGLTGT LTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSIGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN

301 SHEGYGYSDE AVRRHRQGQP *

The originally-identified partial strain B sequence (ORF15) (SEQ ID NO: 78) shows 98.1% identity over a 213aa overlap with ORF15a (SEQ ID NO: 82):

```

5      10      20      30      40      50      60
orf15.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVDMQALHGR
          |||
orf15a     MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVDMQALHGR
          |||
          10      20      30      40      50      60

10     70      80      90     100     110     120
orf15.pep  KVALYIATMGDQSGSLTGGRYSIDAXXGEYINSPAVRTDYTPRYETTAETTSGLTG
          |||
orf15a     KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG
          |||
          70      80      90     100     110     120

15     130     140     150     160     170     180
orf15.pep  LTTSLSLTNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
          |||
orf15a     LTTSLSLTNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
          |||
          130     140     150     160     170     180

20     190     200     210
orf15.pep  FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM
          |||
orf15a     FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
          |||
          190     200     210     220     230     240

```

The complete strain B sequence (ORF15-1) (SEQ ID NO: 80) and ORF15a (SEQ ID NO: 82) show 98.8% identity in 320 aa overlap:

```

30     10      20      30      40      50      60
orf15a.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVDMQALHGR
          |||
orf15-1     MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVDMQALHGR
          |||
          10      20      30      40      50      60

35     70      80      90     100     110     120
orf15a.pep  KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG
          |||
orf15-1     KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG
          |||
          70      80      90     100     110     120

40     130     140     150     160     170     180
orf15a.pep  LTTSLSLTNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
          |||
orf15-1     LTTSLSLTNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
          |||
          130     140     150     160     170     180

45     190     200     210     220     230     240
orf15a.pep  FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
          |||
orf15-1     FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
          |||
          190     200     210     220     230     240

```

		250	260	270	280	290	300
	orf15a.pep	IKPKTNAFEAAAYKENYALWMGPYKVSIGIKPTEGLMVDFSIDIQPYGNHMGNSAPSVEADN					
5	orf15-1	IKPKTNAFEAAAYKENYALWMGPYKVSIGIKPTEGLMVDFSIDIQPYGNHMGNSAPSVEADN					
		250	260	270	280	290	300
		310	320				
	orf15a.pep	SHEGYGYSDEAVRRHRQGPX					
10	orf15-1	SHEGYGYSDEVVRQHRQGPX					
		310	320				

Further work identified the corresponding gene in *N.gonorrhoeae* [<SEQ ID 83>] (SEQ ID NO: 83):

15	1	ATGCGGGCAC	GGCTGCTGAT	ACCTATTCTT	TTTTCAGTTT	TTATTTTATC
	51	CGCCTGCGGG	ACACTGACAG	GTATTCCATC	GCATGGCGGA	GGCAAACGCT
	101	TCGCGGTCGA	ACAAGAACTT	GTGGCCGCTT	CTGCCAGAGC	TGCCGTAA
	151	GACATGGATT	TACAGGCATT	ACACGGACGA	AAAGTTGCAT	TGTACATTGC
	201	AACTATGGGC	GACCAAGGTT	CAGGCAGTTT	GACAGGGGGT	CGCTACTCCA
20	251	TTGATGCACT	GATTTCGCGG	GAATACATAA	ACAGCCCTGC	CGTCCGCACC
	301	GATTACACCT	ATCCGCGTTA	CGAAACCACC	GCTGAAACAA	CATCAGGCGG
	351	TTTGACGGGT	TTAACCACCT	CTTATCTAC	ACTTAATGCC	CCTGCACTCT
	401	CGCGCACCCA	ATCAGACGGT	AGCGGAAGTA	GGAGCAGTCT	GGGCTTAAAT
	451	ATTGGCGGGA	TGGGGGATTA	TCGAAATGAA	ACCTTGACGA	CCAACCCGCG
25	501	CGACACTGCC	TTTCTTTCCC	ACTTGGTGCA	GACCGTATTT	TTCTTGCGCG
	551	GCATAGACGT	TGTTTCTCCT	GCCAATGCCG	ATACAGATGT	GTTTATTAAC
	601	ATCGACGTAT	TCGGAACGAT	ACGCAACAGA	ACCGAAATGC	ACCTATACAA
	651	TGCCGAAACA	CTGAAAGCCC	AAACAAAAC	GGAATATTTT	GCAGTAGACA
	701	GAACCAATAA	AAAATTGCTC	ATCAAACCCA	AAACCAATGC	GTTGAAGCT
	751	GCCTATAAAG	AAAATTACGC	ATTGTGGATG	GGGCCGTATA	AAGTAAGCAA
30	801	AGGAATCAAA	CCGACGGAAG	GATTGATGGT	CGATTTCTCC	GATATCCAAC
	851	CATACGCAA	TCATACGGGT	AACTCCGCC	CATCCGTAGA	GGCTGATAAC
	901	AGTCATGAGG	GGTATGGATA	CAGCGATGAA	GCAGTGCGAC	AACATAGACA
	951	AGGGCAACCT	TGA			

35 This encodes a protein having amino acid sequence [<SEQ ID 84; ORF15ng>] (SEQ ID NO: 84; ORF15ng):

	1	MRARLLIPIL	FSVFILSACG	TLTGIPSHGG	GKRFAVEQEL	VAASARA	AVK
	51	DMDLQALHGR	KVALYIATMG	DQSGSLTGG	RYSIDALIRG	EYINSPAVRT	
40	101	DYTYPRYETT	AETTSGLTGG	LTTSLSTLNA	PALSRTQSDG	SGSRSSLGLN	
	151	IGGMGDYRNE	TLTTNPRDTA	FLSHLVQTVF	FLRGIDVVSP	ANADTDVFIN	
	201	IDVFGTIRNR	TEMHLYNAET	LKAQTKLEYF	AVDRTNKKLL	IKPKTNAFEA	
	251	AYKENYALWM	GPYKVSIGIK	PTEGLMVDFS	DIQPYGNHTG	NSAPSVEADN	
	301	SHEGYGYSDE	AVRQHRQGP	*			

45 The originally-identified partial strain B sequence (ORF15) (SEQ ID NO: 78) shows 97.2% identity over a 213aa overlap with ORF15ng (SEQ ID NO: 84):

orf15.pep	MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARA	AVKMDLQALHGR	60
orf15ng	MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA	AVKMDLQALHGR	60

5	orf15.pep	KVALYIATMGDQSGSLTGGRYSIDAXXXGEYINSPAVRTDYTPRYETTAETTSGGLTG	120
	orf15ng	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG	120
	orf15.pep	LTTSLSLTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF	180
	orf15ng	LTTSLSLTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF	180
	orf15.pep	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM	213
10	orf15ng	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL	240

The complete strain B sequence (ORF15-1) (SEQ ID NO: 80) and ORF15ng (SEQ ID NO: 84) show 98.8% identity in 320 aa overlap:

15	orf15-1.pep	10 20 30 40 50 60 MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
	orf15ng	MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
20	orf15-1.pep	70 80 90 100 110 120 KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG
	orf15ng	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG
25	orf15-1.pep	130 140 150 160 170 180 LTTSLSLTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15ng	LTTSLSLTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
30	orf15-1.pep	190 200 210 220 230 240 FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
	orf15ng	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
35	orf15-1.pep	250 260 270 280 290 300 IKPKTNAFEAAAYKENYALWMGPYKVGKIKPTEGLMVD FSDIRPYGNHTGNSAPSVEADN
	orf15ng	IKPKTNAFEAAAYKENYALWMGPYKVGKIKPTEGLMVD FSDIQPYGNHTGNSAPSVEADN
40	orf15-1.pep	310 320 SHEGYGYSDEVVRQHRQGQPX
	orf15ng	SHEGYGYSDEAVRQHRQGQPX

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (SEQ ID NO: 80) (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 11

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 85>] (SEQ ID NO: 85):

```
1  ..GG.CAGCACA AAAACAGGC GGTGAACGG AAAAACCGTA TTTACGATGA
15  51  TGCCGGGTAT GATATTCGGC GTATTCACGG GCGCATTCTC CGCAAAATAT
    101 ATCCCGCGT TCGGGCTTCA AATTTCTTTC ATCCTGTTTT TAACCGCCGT
    151 CGCATTCAAA ACACTGCATA CCGACCCTCA GACGGCATCC CGCCCGCTGC
    201 CCGGACTGCC CrGACTGACT GCGGTTTCCA CACTGTTCGG CACAATGTCTG
    251 AGCTGGGTCTG GCATAGGCGG CGGTTCACTT TCCGTCCCCT TCTTAATCCA
20  301 CTGCGGCTTC CCCGCCATA AAGCCATCGG CACATCATCC GGCCTTGCCT
    351 GGCCGATTGC ACTCTCCGGC GCAATATCGT ATCTGCTCAA CGGCCTGAAT
    401 ATTGCAGGAT TGCCCGAAGG GTCCTGCGG TCCCTTTACC TGCCCGCCGT
    451 CGCCGTCCTC AGCGCGGCAA CCATTGCCTT TGCCCGGCTC GGTGTCAAAA
    501 CCGCCACAA ACTTTCTTCT GCCAACTCA AAAAATC.TT CGGCATTATG
25  551 TTGCTTTTGA TTGCCGGAAT AATGCTGTAC AACCTGCTTT AA
```

This corresponds to the amino acid sequence [<SEQ ID 86; ORF17>] (SEQ ID NO: 86; ORF17):

```
1  ..GQHKKQAVNG KTVFTMPPGM IFGVFTGAFS AKYIPAFGLQ IFFILFLTAV
30  51  AFKTLHTDPQ TASRPLPGLP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH
    101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPVAV
    151 AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL*
```

Further work revealed the complete nucleotide sequence [<SEQ ID 87>] (SEQ ID NO: 87):

```
35  1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGCGGC
    51  AGGTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
    101 CTGTCGTTTT ATGGGTGCTT GATTTCAGG GTTTGGCACA ACATCCTTAC
    151 GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
    201 CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
    251 CCGTATTTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
```

5
10
301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351 GTTTTTAACC GCCGTGCGAT TCAAAACACT GCATACCGAC CCTCAGACGG
401 CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGACTGCGGT TTCCACACTG
451 TTCGGCACAA TGTCGAGCTG GGTCCGCATA GGCGGCGGTT CACTTTCCGT
501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601 CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCCT
651 TTACCTGCCC GCCGTGCGCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
701 CGCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAAAA
751 Tc.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801 GCTTTAA

This corresponds to the amino acid sequence [<SEQ ID 88; ORF17-1>] (SEQ ID NO: 88; ORF17-1):

15
20
1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTVFTMMP GMIFGVFTGA
101 LSAKYIPAFG LQIFFILELT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL
151 FGTMSWVGI GGGSLVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
251 XFGIMLLLIA GKMLYNLL*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070) (SEQ ID NO: 1119)

25 ORF17 (SEQ ID NO: 86) and HI0902 proteins (SEQ ID NO: 1119) show 28% aa identity in 192 aa overlap:

30
35
ORF17 3 HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
HK + + V + P ++ VF G F + +IF +++L ++ D
HI0902 72 HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130
ORF17 60 QTASRPLPGLPXLTAVSTLFGTMSWVGIGGGSLVPFLIHCGFPAHKAIGTSSGLAWPI 119
Q ++ L L + L G SS GIGGG VPFL G +AIG+S+ +
HI0902 131 QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189
ORF17 120 ALSGAISYLLNGLNIAGLPEGSLGFLYLP AVAVLSAATIAFAPLGVXXXXXXXXXXXXXXX 179
+SG S++++G +PE SLG++YLPVAV ++A + + LG
HI0902 190 GISGMFSFIVSGWGNPLMPEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKG 249
ORF17 180 FGIMLLLIAGKM 191
F + L+++A M
HI0902 250 FALFLIVVAINM 261

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF17 (SEQ ID NO: 86) shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) (SEQ ID NO: 90) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF17a nucleotide sequence [<SEQ ID 89>] (SEQ ID NO: 89) is:

	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAAG	GCAAGTGCAGC
25	51	AGGTTTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTCATGG	TCCTCACCCG
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GAAGTGGAAAA
	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA
30	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCATCCT
	351	GTTTTTAAAC	GCCGTGCGAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG
	451	TTCGGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACCTTCCGT
	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
35	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAAGGGTCAC	TGGGCTTCCT
	651	TTACCTGCCC	GCCGTGCGCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACCGCC	CACAACTTTT	CTTCTGCCAA	ACTCAAAAAA
	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
40	801	GCTTTAA				

This encodes a protein having amino acid sequence [SEQ ID 90] (SEQ ID NO: 90):

45

1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
51	AQHLAGVTSF	AVMVFTAFSS	MLGQHKKQAV	DWKTVFTMMP	GMVFGVFAGA
101	LSAKYIPAFG	LQIFFILFLT	AVAFKTLHTD	PQTASRPLPG	LPGLTAVSTL
151	FGTMSWWGI	GGGSLSPVFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
201	LNGLNIAGLP	EGSLGFLYLP	<u>AVAVLSAATI</u>	<u>AFAPLGVKTA</u>	HKLSSAKLKK
251	SFGIMLLLIA	GKMLYNLL*			

50 ORF17a (SEQ ID NO: 90) and ORF17-1 (SEQ ID NO: 88) show 98.9% identity in 268 aa overlap:

		10	20	30	40	50	60
	orf17a.pep	MWHWDIILILLAVGSAAGFIAGLFGVGGGT	LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF				
5	orf17-1	MWHWDIILILLAVGSAAGFIAGLFGVGGGT	LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF				
		10	20	30	40	50	60
	orf17a.pep	AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMVFGVFAGALS	SAKYIPAFGLQIFFILFLT				
10	orf17-1	AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMIFGVFTGALS	SAKYIPAFGLQIFFILFLT				
		70	80	90	100	110	120
	orf17a.pep	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGG	SLVPFLIHCGFPAHKA				
15	orf17-1	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGG	SLVPFLIHCGFPAHKA				
		130	140	150	160	170	180
	orf17a.pep	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	PAVAVLSAATIAFAPLGVKTA				
20	orf17-1	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	PAVAVLSAATIAFAPLGVKTA				
		190	200	210	220	230	240
	orf17a.pep	HKLSSAKLKKSF	GIMLLLIAGKMLYNLLX				
25	orf17-1	HKLSSAKLKKXFGIMLLLIAGKMLYNLLX					
		250	260	269			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF17 (SEQ ID NO: 86) shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) (SEQ ID NO: 92) from *N. gonorrhoeae*:

30	orf17.pep	GQHKKQAVNGKTVFTMMPGMIFGVFTGAFS	30
	orf17ng	QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKKQAVDWKTI	102
	orf17.pep	AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGTMSSWVGIGG	90
35	orf17ng	AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGG	162
	orf17.pep	GSLSVFPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP	150
	orf17ng	GSLSVFPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLP	202
	orf17.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSF	196
40	orf17ng	AVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKMLYNLL	268

An ORF17ng nucleotide sequence [<SEQ ID 91>] (SEQ ID NO: 91) is predicted to encode a protein having amino acid sequence [<SEQ ID 92>] (SEQ ID NO: 92):

5
1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMIFGVFAGA
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
151 FGAMSSWVG I GGGSLVFPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
251 SFGIMLLLIA GKMLYNLL*

Further work revealed the complete gonococcal DNA sequence [<SEQ ID 93>] (SEQ ID NO: 93):

10
1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCcgtag gcAGTGC GGC
51 AGGTTTTATT GCCGGCCTGT Tcgggtgtagg cggcgGTACG CTGATTGTCC
101 CTGTCGTTTT ATGGGTGCTT GATTTGCAGG GTTTGGCACA ACATCCTTAC
151 GCGCAACACC TCGCCGTCGG CACA Tccttc gcCGTCATGG TCTTCACCGC
201 CTTTTCAGT ATGTTGGGGC AGCACAAAA ACAGGCGGTC GACTGGAAAA
251 CCATATTTGC GATGATGCCG GGTATGATAT TCGGCGTATT CGCTGGCGCA
15
301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351 GTTTTAAACC GCCGTCGCAT TCAAAACACT GCATACCGGT CGTCAGACGG
401 CATCCCGCCC GCTGCCCGGG CTGCCCGGAC TGACTGCGGT TTCCACACTG
451 TTCGGCGCAA TGTCGAGCTG GGTCCGCATA GGCGGCGGTT CACTTTCCGT
20
501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGC GCAAT ATCGTATCTG
601 GTCAACGGTC TGAATATTGC AGGATTGCCC GAAGGGTCGC TGGGCTTCCT
651 TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCC
701 CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAGAA
25
751 TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801 GCTTTAA

This corresponds to the amino acid sequence [<SEQ ID 94; ORF17ng-1>] (SEQ ID NO: 94;
ORF17ng-1):

30
1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMIFGVFAGA
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
151 FGAMSSWVG I GGGSLVFPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
35
251 SFGIMLLLIA GKMLYNLL*

ORF17ng-1 (SEQ ID NO: 94) and ORF17-1 (SEQ ID NO: 88) show 96.6% identity in 268 aa overlap:

40
orf17-1.pep 10 20 30 40 50 60
MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
|||||
orf17ng-1 MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
10 20 30 40 50 60
45
orf17-1.pep 70 80 90 100 110 120
AVMVFTAFSSMLGQHKQAVDWKTIVFTMPGMIFGVFTGALS AKYIPAFGLQIFFILFLT
|||||
orf17ng-1 AVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVFAGALS AKYIPAFGLQIFFILFLT
70 80 90 100 110 120
50
orf17-1.pep 130 140 150 160 170 180
AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLVFPFLIHCGFPAHKA
|||||

-135-

5 orf17ng-1 AVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGGGSLVPFLIHCGFPAHKA
130 140 150 160 170 180

orf17-1.pep IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA
190 200 210 220 230 240

orf17ng-1 IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA
190 200 210 220 230 240

10 orf17-1.pep HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
250 260 269

orf17ng-1 HKLSSAKLKESFGIMLLLIAGKMLYNLLX
250 260

In addition, ORF17ng-1 (SEQ ID NO: 94) shows significant homology with a hypothetical
15 *H.influenzae* protein (SEQ ID NO: 1119):

sp|P44070|Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir|G64015 hypothetical protein
HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae
predicted coding region HI0902 [Haemophilus influenzae]Length = 264
Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
Identities = 15/43 (34%), Positives = 23/43 (53%)

Query: 55 AVGTSFAVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVF 97
A+GTSFA +V T S HK + W+ + + P ++ VF
Sbjct: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94

Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
Identities = 44/114 (38%), Positives = 65/114 (57%)

Query: 150 LFGAMSSWVGIGGGSLVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209
L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G +
Sbjct: 148 LIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207

Query: 210 PEGSLGFLYLPAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263
PE SLG++YLPVAV ++A + + LG KL + LK+ F + L+++A M
Sbjct: 208 PEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261

This analysis, including the homology with the hypothetical *H.influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 12

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 95>] (SEQ ID NO: 95):

1 ..GGAAACGGAT GGCAGGCAGA CCCCGAACAT CCGCTGCTCG GGCTTTTTGC

51 CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCGGAATA TGTGCGTTGG
 101 TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTTTGTGTT TGCGGCACTG
 151 CTCAAACTTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
 201 GCTGATGGCG GTTGCCATATGCCACCGCTG CGGTATAGAC CGGCAGCCGC
 5 251 CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGGCAGCG
 301 TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA
 351 A

This corresponds to the amino acid sequence [<SEQ ID 96; ORF18>] (SEQ ID NO: 96; ORF18):

10 1 ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
 51 LKLYALKPVY WFLVQFVLMV VAYVHRCGID RQPPSTFGGS QLRLGGLTAA
 101 LMQVSVLVLL LSEIGR*

Further work revealed the complete nucleotide sequence [<SEQ ID 97>] (SEQ ID NO: 97):

15 1 ATGATTTTGC TGCATTGGA TTTTGTGCT GCCTTACTGT ATGCGGCGGT
 51 TTTTCTGTTT CTGATATTC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
 101 GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
 151 GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTTCATCC CCCATTTTTTA
 201 CCTGACTTTG GGCAGCATAT TTTTTCATCG GGGCATTGG AACCGGAAAA
 20 251 CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GTCGGGCTT
 301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTGTGCG GAATATGTGC
 351 GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTT GTGTTTGGCG
 401 CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTGTCAG
 451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
 25 501 GCCGCCGTCA ACGTTCGGCG GCTCGCAGCT GCGACTCGGC GGGTTGACGG
 551 CAGCGTTGAT GCAGGTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
 601 AGATAA

This corresponds to the amino acid sequence [<SEQ ID 98; ORF18-1>] (SEQ ID NO: 98; ORF18-1):

30 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
 51 GIWGMTRAAP LFIPHFYLTLSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
 101 FAVSNVSM TLAFVICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
 151 FVLMVAVYVH RCGIDRQPPS TFGGSQRLRG GLTAALMQVS VLVLLSEIG
 201 R*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF18 (SEQ ID NO: 96) shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) (SEQ ID NO: 100) from strain A of *N. meningitidis*:

40 orf18.pep
 orf18a
 10 20 30
 GNGWQADPEHPLLGLFAVSNVSM TLAFVGI
 |||||
 TRAAPLFIPHFYLTLSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGI
 60 70 80 90 100 110

-137-

5

	40	50	60	70	80	90
orf18.pep	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAYVHRCGIDRQPPSTFGGS					
orf18a	CALVHYCFSTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAYVHRCGIDRQPPSTFGGS					
	120	130	140	150	160	170

	100	110
orf18.pep	QLRLGGLTAALMQVSVLVLLLSEIGRX	
orf18a	QLRLGGLTAALMQXSVLVLLLSEIGRX	
	180	190 200

10

The complete length ORF18a nucleotide sequence [<SEQ ID 99>] (SEQ ID NO: 99) is:

15

1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCCC
151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TTGTTTCATCC	CCCATTTTTTA
201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGCATTGG	AACCGGAAAA
251	CGGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCTCT	GCTCGGGCTG
301	TTTGCCGTCA	GTAATGTATC	GATGACGCTT	GCTTTTGTCT	GAATATGTGC
351	GTTGGTGCA	TATTGCTTTT	CGNGAACGGT	TCAAGTGTTT	GTGTTTGCGG
401	CACTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCAG
451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTGCGGTA	TAGACCGCA
501	GCCGCCGTCA	ACGTTTCGGC	GNTCGCAGCT	GCGACTCGGC	GGGTTGACGG
551	CAGCGTTGAT	GCAGNTCTCG	GTACTGGTGC	TGCTGCTTTC	AGAAATTGGA
601	AGATAA				

20

25

This encodes a protein having amino acid sequence [<SEQ ID 100>] (SEQ ID NO: 100):

30

1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP
51	GIWGMTRAAP	LFIPHFYLT	LSIFFFIGHW	NRKTDGNGWQ	ADPEHPLLGL
101	FAVSNVSM	TLAFV	GICALVH	YCF	SXTVQVF
151	FVLMAYAYVH	RCGIDRQPPS	TFGGSQRLRG	GLTAALMQXS	VLVLLLSEIG
201	R*				

ORF18a (SEQ ID NO: 100) and ORF18-1 (SEQ ID NO: 98) show 99.0% identity in 201 aa overlap:

40

	10	20	30	40	50	60
orf18a.pep	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
orf18-1	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
	10	20	30	40	50	60

	70	80	90	100	110	120
orf18a.pep	LFIPHFYLTLSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM					
orf18-1	LFIPHFYLTLSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM					
	70	80	90	100	110	120

	130	140	150	160	170	180
orf18a.pep	YCFSTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAYVHRCGIDRQPPSTFGGSQRLRG					
orf18-1	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAYVHRCGIDRQPPSTFGGSQRLRG					
	130	140	150	160	170	180

50

```

                                190      200
orf18a.pep  GLTAALMQSVLVLLSEIGRX
              |||||
orf18-1     GLTAALMQSVLVLLSEIGRX
              190      200

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 (SEQ ID NO: 96) shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) (SEQ ID NO: 102) from *N. gonorrhoeae*:

10	orf18.pep		GNGWQADPEHPLLGLFAVSNVSMTLAFVGI	30
	orf18ng	TRAAPLFIPHFYLTLGSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI		115
	orf18.pep	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGS		90
	orf18ng	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGS		175
15	orf18.pep	QLRLGGLTAALMQVSVLVLLSEIGR	116	
		: : : :: :		
	orf18ng	QLRLGVLAAMLQVAVTAMLLAEIGR	201	

The complete length ORF18ng nucleotide sequence is [<SEQ ID 101>] (SEQ ID NO: 101):

20	1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGt	aTGCGGcggt
	51	tttTctgTtT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTGCGTT	GTGGCTCGGC	ATCTCGGTTT	TAGGGGTAAA	GCTGATGCCG
	151	GGGATGTGGG	GAATGACCCG	CGCCCGCGCT	TTGTTTCATCC	CCCATTTTTTA
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGTATTGG	AACCGGAAAA
25	251	CAGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
	301	TTTGCCGTCa	GTAATGTATC	GATGACGCTT	GCTTTTGTCTG	GAATATGTGC
	351	GTTGGTGCAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT	GTGTTTGC GG
	401	CATTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCAG
	451	TTTGTATTGA	TGGCGGttgC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
30	501	GCCGCCGTCA	ACGTTCCGGC	GTTCGCAGCT	GCGACTCGGC	GTGTTGGCGG
	551	CGATGTTGAT	GCAGGTTGCG	GTAACGGCGA	TGCTGCTTGC	CGAAATCGGC
	601	AGATGA				

This encodes a protein having amino acid sequence [<SEQ ID 102>] (SEQ ID NO: 102):

35 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLG ISVLGVKLMP
 51 GMWGMTRAAP LFIPHFYLT GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL
 101 FAVSNVSMTL AFVGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
 151 FVLMAVAVH RCGIDRQPPS TFGGSQLRLG VLAAMLQVA VTAMLLAEIG
40 201 R*

This ORF18ng (SEQ ID NO: 102) protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1 (SEQ ID NO: 98):

orf18-1.pep 10 20 30 40 50 60
MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP

	orf18ng	MILLHLDFLSALLYAAVFLFLIFRAGMLQFWASIALWLGISVLGVKLMPGMWGMTRAAP
		10 20 30 40 50 60
5	orf18-1.pep	LFIPHFYLTLSGSIFFFIHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
	orf18ng	LFIPHFYLTLSGSIFFFIHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
		70 80 90 100 110 120
10	orf18-1.pep	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGSQRLRG
	orf18ng	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGSQRLRG
		130 140 150 160 170 180
15	orf18-1.pep	GLTAALMQVSVLVLVLLSEIGRX
	orf18ng	VLAAMLQVAVTAMLLAEIGRX
		190 200

Based on this analysis, including the presence of several putative transmembrane domains in the
 20 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 13

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 103>] (SEQ ID
 NO: 103):

25	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTN	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTT	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTTGTCGAT
	151	TTGGACAACC	NCNTGACCGG	ACGGCTNAAA	AACATCATCA	CCACGTCGC
	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCACA	AAGCACCTC	GGCACAGGGC
30	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CTT.CG.CTT	CACCATTTTA
	301	GGCGCGNCG	...			

This corresponds to the amino acid sequence [<SEQ ID 104; ORF19>] (SEQ ID NO: 104;
 ORF19):

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD
	51	LDNXXTGRLK	NIITTVLFT	LSSLTAQSTL	GTGLPFILAM	TLMTXXFTIL
	101	GAX...				

Further work revealed the complete nucleotide sequence [<SEQ ID 105>] (SEQ ID NO: 105):

40	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTT	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA

5 101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTGCGAT
 151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC
 201 CCTGTTCAACC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTTA
 301 GGC GCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
 401 ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
 451 CTGTTCCAAA TCGTCCTGCC CCACGCCCC GTCCAAGAAA GCGTCGCCAA
 501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
 10 551 ACCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
 651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
 701 GTTACTACTT TGCCGCCCAA GACATACACG AACGCATCAG CTCGCCCCAC
 751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
 15 801 CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
 851 CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
 901 CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
 951 CGACAGTCCC GACATCCGCC ACCTGCGCGG CTTTCTCGAC AACCTCGGCA
 20 1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
 1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
 1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
 1151 TATTCGCGCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
 1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
 25 1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
 1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
 1351 TACTTCACCC CGTCTGTGCA AACCAAATC TGGATTGTCA TCGCCAGTAC
 1401 CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
 1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
 30 1501 TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCCT
 1551 TGCCTGGCG CGAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
 1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
 1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
 35 1751 CCCTTTCGCA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
 1801 CCCGCTTTA CCCTGCTCAA AACCGGTAC GCCCTGACCG GCTACATCTC
 1851 CGCCCTCGGC GCATACGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
 1901 TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
 1951 CACCTGCCCC AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
 40 2001 GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGCT CGAACCCCTAC
 2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
 2151 A

This corresponds to the amino acid sequence [<SEQ ID 106; ORF19-1>] (SEQ ID NO: 106;

45 ORF19-1):

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGR LK NIITVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
 151 LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
 50 201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFAAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIFRIH RLLEMQQQAC RNTAQALRAS KDYVYSKRLG
 301 RAIEGCRQSL RLLSDSNDSP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
 351 NDRMGDTRIA ALETSSLKNT WQAI RPQLNL ESGVFRHAVR LSLVVAAACT
 401 IVEALN LNLG YWILLTALFV CQPNYTATKS RVRQR IAGTV LGVI VGS LVP
 55 451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
 551 KITERLKSGE TGDDVEYRAT RRRRAHEHTAA LSSTLSDMSS EPAKFADSLQ
 601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ

651 HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701 YRAYROIPIHR OPONAA*

Computer analysis of this amino acid sequence gave the following results:

- 5 Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289) (SEO ID NO: 1120)

ORF19 (SEQ ID NO: 104) and YHFK proteins (SEQ ID NO: 1120) show 45% aa identity in 97 aa overlap:

10

orf19	6	LKPLLITSLP VFASVF TAASIVWQLGEPKLAMPFVLGIIAGGLVLDLDNXXTGR LKNIITT	65
		L +I+++PVF +V AA +W +MP +LGIIAGGLVLDLDN TGR LKN+ T	
YHFK	5	LNAKVISTIPVFIAVNIAAVGIWFFDISQSMP LILGIIAGGLVLDLDNRLTGR LKNVFFT	64
orf19	66	VALFTLSSLTAQSTLGTGLPFILAMTLM TXXFTILGA	102
		+ F++SS Q +G + +I+ MT++T FT++GA	
YHFK	65	LIAFSSISFIVOLHIGKPIQYIVLMTVLTFFIPTMIGA	101

- 15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 (SEQ ID NO: 104) shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) (SEQ ID NO: 108) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
20	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK					
	orf19a	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK					
		10	20	30	40	50	60
		70	80	90	100		
25	orf19.pep	NIITTVALFTLSSSLTAQSTLGTGLPFI ILAMTLMTXXTILGAX					
	orf19a	NIIATVALFTLSSSLVAQSTLGTGLPFI ILAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
	orf19a	TTLTYTPETYWLTNPF MILCGTVLYSTAILFQI ILPHRPVQENVANAYEALGSYLEAKA					
		130	140	150	160	170	180

The complete length ORF19a nucleotide sequence [<SEQ ID 107>] (SEQ ID NO: 107) is:

	1	ATGAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTG	GGCGAACCCA
35	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGTGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTCAAC	CTCTCCTCAC	TTGTCGCGCA	AAGCACCCTC	GGCACAGGTT
	251	TGCCATTCA	CCTCGCCATG	ACCTGATGA	CTTTCGGCTT	TACCATCATG
	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACCTCGCCGT
	351	CGCCACCTAC	ACCACACTTA	CCTACACCCC	CGAAACCTAC	TGGGTGACCA
40	401	ACCCCTTTAT	GATTCTGTGC	GGAACCGTAC	TGTACAGCAC	CGCCATCATC

5
10
15
20
25
30
35

```

451 CTGTTCCAAA TCATCCTGCC CCACCGCCCC GTTCAAGAAA ACGTCGCCAA
501 CGCCTACGAA GCACTCGGCA GCTACCTCGA AGCCAAAGCC GACTTTTTCG
551 ATCCCGACGA AGCGGAATGG ATAGGCAACC GCCACATCGA CCTCGCCATG
601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
701 GCTACTACTT CGCGGCCCAA GACATACACG AACGCATCAG CTCCGCCAC
751 GTCGACTACC AAGAGATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
801 CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
851 CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
901 CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
951 CGACAATCCC GACATCCGCC ACCTGCGCCG CCTTCTCGAC AACCTCGGCA
1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
1051 AACGACCGCA TGGCGGACAC CCGCATCGCC GCCCTCGAAA CCGGCAGCCT
1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTTG TCGTTGCCGC CGCCTGCACC
1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAGC CGCGTCCGCC
1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTTACCC CCTCCGTCGA AACCAAATC TGGATCGTCA TCGCCAGTAC
1401 CACCTCTTTT TTCATGACCC GCACCTACAA ATACAGCTTC TCGACATTTT
1451 TCATCACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG GTTGGACGTA
1501 TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
1551 TGCTTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGCGC CTATCTCGAA
1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTGCGCGA CAGCCTGCAA
1801 CCCGGCTTTA CCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCC AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
2001 GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGGCAGCT CGAACCCTAC
2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151 A

```

This encodes a protein having amino acid sequence [SEQ ID 108] (SEQ ID NO: 108):

40
45
50

```

1 MKTPPLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNRLTGRLK NIIATVALFT LSSLVAQSTL GTGLPFILAM TLMTFGFTIM
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI
151 LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG
301 RAIEGCRQSL RLLSDSNDNP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
351 NDRMGDTRIA ALETGSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAACT
401 IVEALNLNLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQUAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
551 KITERLKSGE TGDDVEYRAT RRRRAHEHTAA LSSTLSDMSS EPAKPADSLQ
601 PGFTLLKGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPETEPDDF QTALDTRLGE LDTLRTHSSG TQSHILLOQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*

```

ORF19a (SEQ ID NO: 108) and ORF19-1 (SEQ ID NO: 106) show 98.3% identity in 716 aa overlap:

5	orf19a.pep	10 20 30 40 50 60 MKT PPLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
	orf19-1	10 20 30 40 50 60 MKT PPLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
10	orf19a.pep	70 80 90 100 110 120 NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY : : : :
	orf19-1	70 80 90 100 110 120 NIIITVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY : : : :
15	orf19a.pep	130 140 150 160 170 180 TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA : : : :
	orf19-1	130 140 150 160 170 180 TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIVLPHRPVQESVANAYDALGGYLEAKA : : : :
20	orf19a.pep	190 200 210 220 230 240 DFFDPDEAEWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
	orf19-1	190 200 210 220 230 240 DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
25	orf19a.pep	250 260 270 280 290 300 DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG
	orf19-1	250 260 270 280 290 300 DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG
30	orf19a.pep	310 320 330 340 350 360 RAIEGCRQSLRLSDSNDNPDIRHLRRLLDNLGSVDQQFROLQHNGLQAENDRMGDTRIA
	orf19-1	310 320 330 340 350 360 RAIEGCRQSLRLSDSNDSPDIRHLRRLLDNLGSVDQQFROLQHNGLQAENDRMGDTRIA
35	orf19a.pep	370 380 390 400 410 420 ALETGSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAACTIVEALNNLGYWILLTALFV :
	orf19-1	370 380 390 400 410 420 ALETSSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAACTIVEALNNLGYWILLTALFV :
40	orf19a.pep	430 440 450 460 470 480 CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
	orf19-1	430 440 450 460 470 480 CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
45	orf19a.pep	490 500 510 520 530 540 STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIIGASLAWAAVSYLWPDWKYLTALERTAL
	orf19-1	490 500 510 520 530 540 STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIIGASLAWAAVSYLWPDWKYLTALERTAL
	orf19a.pep	550 560 570 580 590 600 AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSSTLSDMSSEPAKFADSLQ

[illegible]

Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF19 (SEQ ID NO: 104) shows 95.1% identity over a 102aa overlap with a predicted ORF
(ORF19.ng) (SEQ ID NO: 110) from *N. gonorrhoeae*:

	orf19.pep	MKTPLLKPLLITSLPVFASVF TAASIV WQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVF TAASIV WQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK	60
20	orf19.pep	NIITTVALFTLSSSLTAQSTLGTGLPFILAMTLM TX FTILGAX	103
	orf19ng	NIITATVALFTLSSSLTAQSTLGTGLPFILAMTLM TF GFTILGAVGLKYRTFAFGALAVATY	120

25 An ORF19ng nucleotide sequence [<SEQ ID 109>] (SEQ ID NO: 109) is predicted to encode a protein having amino acid sequence [<SEQ ID 110>] (SEQ ID NO: 110):

30

1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIAGGLVD
51	<u>LDNRLTGRLK</u>	<u>NIIATVALFT</u>	<u>LSSLTAQSTL</u>	GTGLPFILAM	TLMTFGFTIL
101	GAVGLKYRTF	AFGALAVATY	<u>TTLTYPETY</u>	<u>WLTNPFMILC</u>	GTVLYSTAII
151	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA	DDFDPEAAW	IGNRHIDLAM
201	SNTGVITAFN	QCRSALFYRL	RGKHRHPTA	KMLRYFYAAQ	DIHERISSAH
251	VYQEMSEKF	KNTDIIFRIR	RLLEMQGQAC	RNTAQAIRSG	KDYVYSKRLG
301	RAIEGCRQSL	RLSDGNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHSDSPAE
351	NDRMGDTRIA	ALETGSFKNT	*		

35 Further work revealed the complete nucleotide sequence [<SEQ ID 111>] (SEQ ID NO: 111):

	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTGGTCGAT
40	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCGCA	AAGCACCCTC	GGCAGAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CCTCGCCCGT
	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
45	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	TGCCAAGAAA	CGCGTGCCAA
	501	TGCCTACGAA	GCACCTCGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG

5 551 ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
651 TTACCGTTTG CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
701 GCTACTACTT CGCCGCCCAA GACATCCACG AACGCATCAG CTCCGCCAC
751 GTCGACTACC AAGAGATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
801 CCGCATCCGC CGCCTGCTCG AAATGCAGGG GCAGGCGTGC CGCAACACCG
851 CCAAGCCAT CCGGTCGGGC AAAGACTAcg tTTACAGCAA ACGCCTCGGA
901 CGCGCCATcg aaggctgCCG CCAGTCGctg cgcctCCTTt cagacggcaA
951 CGACAGTCCC GACATCCGCC ACCTGAGccg CCTTCTCGAC AACCTCGgca
10 1001 GCGTcgacca gcagtTCcg caactCCGAC ACAGcgactC CCCCgcgaa
1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCCtcgaaa ccggcagctT
1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC
15 1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC
1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CCTCCGTCGA AACCAACTC TGGATTGTCA TCGCCGTAC
1401 CACCCTGTTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
20 1501 TACGCCGCCA TGCCCGTGCG CATCATcgaC ACCATTATCG GCGCATCCCT
1551 TGCCTGGCG GCGGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA
1651 AAAATTGCCG AACGCCTCAA AACCGCGGAA ACCGGCGACG ACATAGAATA
1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
25 1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
1801 CCCGGCTTTA CCCTGCTCAA AACCGGTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCGGACT
1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACGCCCA CATCTTCCAA
1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT
30 2001 GCGCGGCGAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC
2051 ACATCCTCCT CCAACGCTC CAACTCATCG CccgGCAACT CGAACCTTAC
2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151 A

35 This corresponds to the amino acid sequence [<SEQ ID 112; ORF19ng-1>] (SEQ ID NO: 112;
ORF19ng-1):

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNRLTGR LK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFILIC GTVLYSTAI I
40 151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIR RLLEMQQQAC RNTAQAIRSG KDYVYSKRLG
301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPA E
351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAACT
45 401 IVEALNLNLG YWILLTALFV CQPNYTATKS RVYQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSSGTYLQ
551 KIAERLKTGE TGDDIEYRIT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
50 651 HLPDMGPDDF QTALDTLRGE LGTLRTRSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*

ORF19ng-1 (SEQ ID NO: 112) and ORF19-1 (SEQ ID NO: 106) show 95.5% identity in 716 aa overlap:

		10	20	30	40	50	60
	orf19-1.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK					
5	orf19ng-1	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK					
		10	20	30	40	50	60
	orf19-1.pep	NIITTVALFTLSSLTAQSTLTGLPFIAMTLMTFGFTILGAVGLKYRTFAFGALAVATY					
10	orf19ng-1	NIITATVALFTLSSLTAQSTLTGLPFIAMTLMTFGFTILGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
	orf19-1.pep	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA					
15	orf19ng-1	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIILPHRPVQESVANAYEALGGYLEAKA					
		130	140	150	160	170	180
	orf19-1.pep	DFPDPDEAAWIGNRHIDLAMSTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ					
20	orf19ng-1	DFPDPDEAAWIGNRHIDLAMSTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ					
		190	200	210	220	230	240
	orf19-1.pep	DIHERISSAHVDYQEMSEKFKNTDIIFRIRLLEMQGQACRNTAQALRASKDYVYSKRLG					
25	orf19ng-1	DIHERISSAHVDYQEMSEKFKNTDIIFRIRLLEMQGQACRNTAQAIRSGKDYVYSKRLG					
		250	260	270	280	290	300
	orf19-1.pep	RAIEGCRQSLRLLSDSNDSPDIRHLRLLDNLGSDVQQFRQLQHNGLOAENDRMGDTRIA					
30	orf19ng-1	RAIEGCRQSLRLLSDGNDSPDIRHLRLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIA					
		310	320	330	340	350	360
	orf19-1.pep	ALETSSLKNTWQAIROPQNLNLESGVFRHAVRLSLVAAACTIVEALNRLNGYWILLTALFV					
35	orf19ng-1	ALETGSFKNTWQAIROPQNLNLESCVFRHAVRLSLVAAACTIVEALNRLNGYWILLTALFV					
		370	380	390	400	410	420
	orf19-1.pep	CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF					
40	orf19ng-1	CQPNYTATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF					
		430	440	450	460	470	480
	orf19-1.pep	STFFITIQAALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAAL					
45	orf19ng-1	STFFITIQAALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLERTAAL					
		490	500	510	520	530	540
	orf19-1.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ					
		550	560	570	580	590	600

Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 113>] (SEQ ID NO: 113):

```

5      1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
      51  GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
     101  CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAAC T GCCCAACCTG
     151  CTTCCGCGCG TGTTCGCGGA GGGGCGGTTT GCCCAAGCGT TTGTGCCGAT
     201  TTTGGCGGAA TACAAGGAAA CGCGTTCAA AGAGGCGG.C GAAGCCTTTA
     251  TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
     301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
     351  TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATT T GCTGCGGATT
     401  ACGTTTCCTT ATATATTATT GATTTCCTTG TCTTCATTG TCGGCTCGGT
     451  ACTCAATTCT TATCATAAGT TCGGCATTCC GGCGTTTACG CCAC.GTTTC
     501  TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTCGTGCC GTATTTTCGAT
     551  CCGCCCGTTA CCGCGCGyGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGCA
     601  ACTCGrmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
     651  CCAAACtGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
     701  GCGCTGCGa TTTTgGGCGT GAgCGTGGCG CAGGTTTCTT TGGTGATCAA
     751  CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
     801  ACGCCGACCG CATGATGGAG CTGCCCAGCG GCGTGCTGGG GGCGGCACTC
     851  GTTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC
     901  GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGctgc
     951  TGACGCTGCC GGCgGcGGTC GGACTGGCGG TGTGTGCTT cCCgCtGGTG
    1001  GCGACGCTGT TTATGTACCG CGwATTTACG CTGTTTGACG CGCAGATGAC
    1051  GCAACACGCG CTGATTGCCT ATTCTTTCGG TTTAATCGGC TTAATCATGA
    1101  TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAwAmGCCC
    1151  GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
    1201  CTTTAYCGGC CCACTrrAAC rCaGTCGGAC TTTCGCTTGC CATCGGTCTG
    1251  GGCgCGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTTGC GCAGACACGG
    1301  TATTTACCAA CCTGG.CAAG GGTGGGCGAG CGTTCTT.AG CAAAATGCT
    1351  GcTCTCGCTC GCCGTGA

```

This corresponds to the amino acid sequence [<SEQ ID 114; ORF20>] (SEQ ID NO: 114; ORF20):

```

35      1  MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
      51  LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
     101  LGILAAPWVI YVSAPSFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
     151  LNSYHKFGIP AFTPXFLNVS FIVFALFFVP YFDPPVTAXA WAVFVGILQ
     201  LXFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
     251  TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
     301  EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
     351  QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
     401  FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
     451  SRSP*

```

These sequences were elaborated, and the complete DNA sequence [<SEQ ID 115>] (SEQ ID NO: 115) is:

```

1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
51  GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG

```

5
 10
 15
 20
 25
 30

```

101  CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACCT GCCCAACCTG
151  CTTCCGCCGCG TGTTTGCGGA GGGGGCGGTTT GCCCAAGCGT TTGTGCCGAT
201  TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGGCG GAGGCTTTTAA
251  TCCGCCATGT GGC GGCGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
351  TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTTG CTGCGGATTA
401  CGTTTCCTTA TATATTATTG ATTTCCCTGT CTTCAATTTGT CGGCTCGGTA
451  CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT
501  GAACGTGTCTG TTTATCGTAT TCGCGCTGTT TTTGCTGCCG TATTCGATC
551  CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTCGGCGG CATTTTGCAA
601  CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
651  CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
701  CGCCTCGGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
751  ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
801  CGCCGACCGC ATGATGGAGC TGCCCAGCGG CGTGCTGGGG GCGGCACTCG
851  GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
901  GAACAGTTTT CCGCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
951  GACGCTGCCG GCGGCGGTCT GACTGGCGGT GTTGTGTTT CCGCTGGTGG
1001 CGACGCTGTT TATGTACCGC GAATTTACGC TGTTTGACGC GCAGATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT
1101 TAAAGTGTTG GCACCCGGCT TCTATGCGCG GCAAAACATC AAAACGCCCG
1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC
1201 TTTATCGGCC CACTGAAACA CGTCGGAATT TCGCTTGCCA TCGTCTGGG
1251 CGCGTGATAT AATGCCGGAT TGTGTTTATA CCTGTGCGC AGACACGGTA
1301 TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTAGCAAA AATGCTGCTC
1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTACCTGCC
1401 GTTTGAATGG GCGCAGCCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GGCGGCTTTG
1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAACTGA
  
```

This corresponds to the amino acid sequence [SEQ ID 116; ORF20-1] (SEQ ID NO: 116; ORF20-1):

35
 40

```

1  MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFVAFKLPNL
51  LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA
101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151 LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPPTVATA WAVFVGGILQ
201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
501 GFRPRHFKRV EN*
  
```

45 Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169) (SEQ ID NO: 1122)

ORF20 (SEQ ID NO: 114) and MviN proteins (SEQ ID NO: 1122) show 63% aa identity in 440aa overlap:

-150-

	Orf20	1	MNMLGALAKVGS	10	TMVSRVLGFVRD	20	TVIARAFGAGMAT	30	DAFFVAFKLPNLLRRVFAEGAF	40	60
	MviN	14	MNLLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF	20		30		40		50	73
5	Orf20	61	AQAFVPILA	70	EYKETRSKEAXEAFIRHVAGMLS	80	SVFLVIVTALGILAAPWVIYVSAPSFAQD	90	120		
	MviN	74	SQAFVPILA	80	EYKSKQGEETRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT	90	133				
10	Orf20	121	ADKFQLSIDLLRITFPYILLISLSS	130	FVGSVLNSYHKFGIPAFTPXFLNV	140	SFIVFALFFVP	150	180		
	MviN	134	ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP	140	193						
15	Orf20	181	YFDPPTVTA	190	XAVFVGGILQLXFLPWLAKLGLKLPKLSFKDAAVNRVMQMAPAILGV	200	240				
	MviN	194	YFNPPVLALAWAVTVGGVLQVYQLPYLKKIGMLVLP	200	253						
20	Orf20	241	SVAQVSLVINTIFASYLQSGSVSWMY	250	YADRMELPSGVLGAALGTILLPTLSKHSANQDT	260	300				
	MviN	254	SVSQISLIINTIFASFLASGSVSWMYADRLMEFPSGVLGVALGTILLPSLSKSFASGNH	260	313						
25	Orf20	301	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTL	310	FDAQMTQHAIAYSFG	320	360				
	MviN	314	DEYCRIMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG	320	373						
30	Orf20	361	LIGLIMIKVLAPGFYARQNI	370	XPVKIAIFTLICXQLMNLXFX	380	XXXXXXXXXXXXXXXXXXXXCI	390	420		
	MviN	374	LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL	380	433						
35	Orf20	421	NAGLLFYLLRRHGIQ	430	PXQG	440					
	MviN	434	NASLLYWQLRKQNI	440	FTPQPG	453					

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 (SEQ ID NO: 114) shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) (SEQ ID NO: 118) from strain A of *N. meningitidis*:

30	orf20.pep	10	20	30	40	50	60
	orf20a	10	20	30	40	50	60
35	orf20.pep	70	80	90	100	110	120
	orf20a	70	80	90	100	110	120
40	orf20.pep	130	140	150	160	170	180
	orf20a	130	140	150	160	170	180
45	orf20.pep	190	200	210	220	230	240
	orf20a	190	200	210	220	230	240

orf20.pep	<u>YFDPPVTAXAWAVFVGIGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV</u>
orf20a	<u>YFDPPVTALAWAVFVGIGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV</u>
	190 200 210 220 230 240
orf20.pep	250 260 270 280 290 300
orf20a	<u>SVAQVSLVINTIFASYLQSGSVSWMYADRMMEPSGVLGAAAGTILLPTLSKHSANQDT</u>
	250 260 270 280 290 300
orf20.pep	310 320 330 340 350 360
orf20a	<u>EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG</u>
	310 320 330 340 350 360
orf20.pep	370 380 390 400 410 420
orf20a	<u>LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI</u>
	370 380 390 400 410 420
orf20.pep	430 440 450
orf20a	<u>NAGLLFYLLRRHGIYQXPQGLGSVLXQKCCSRSPX</u>
	430 440 450 460 470 480

25 The complete length ORF20a nucleotide sequence [<SEQ ID 117>] (SEQ ID NO: 117) is:

	1	ATGAATATGC	TGGGAGCTTT	GGTAAAAGTC	GGCAGCCTGA	CGATGGTGTCT
	51	GCGCGTTTTG	GGATTTGTGC	GCGATACGGT	CATTGCGCGC	GCATTCGGCG
	101	CAGGCATGCG	GACGGATGCG	TTCTTTGTCTG	CGTTCAAACCT	GCCCAACCTG
30	151	CTTCGCCGCG	TGTTTGCGGA	GGGGGCGTTT	GCCCAAGCGT	TTGTGCCGAT
	201	TTTGGCGGAA	TATAAGGAAA	CGCGTCTCTAA	AGAGGCGACG	GAGGCTTTTA
	251	TCCGCCATGT	GGCGGGGATG	CTGTCTGTTT	TACTGGTTCAT	CGTTACCGCG
	301	CTGGGCATAC	TTGCCGCGCC	TTGGGTGATT	TATGTTTCCG	CACCCGGTTT
	351	TGCCAAAGAT	GCCGACAAAT	TTCAGTCTCT	TATCGATTTC	CTGCGGATTA
35	401	CGTTTCCTTA	TATCTTATTG	ATTTCACTTT	CCTCTTTTGT	CGGCTCGGTA
	451	CTCAATTCCCT	ATCATAAAAT	CAGCATTCCT	CGGTTTACGC	CCACGTTCTT
	501	GAACGTGTGC	TTTATCGTAT	TGCGCTGTTT	TTTCGTCGCG	TATTTGATAT
	551	CTCCCGTTAG	CGCGCTGGCT	TGGGCGGTTT	TTGTGCGCGG	CATTTTGCAA
	601	CTCGGCTTCC	AACTGCCCTG	GCTGGCGAAA	CTGGGTTTTT	TGAAACTGCC
40	651	CAAACGTAGT	TTCAAAGATG	CGGCGGTCAA	CCGCGTGATG	AAACAGATGG
	701	CGCCTGCGAT	TTTGGGCGTG	AGCGTGCGCG	AGATTTCTTT	GGTGATCAAC
	751	ACGATTTTCG	CGTCTTATCT	GCAATCGGGC	AGCGTTTCTA	GGATGTATTA
	801	CGCCGACCGC	ATGATGGAAC	TGCCCGCGCG	CGTGCTGGGG	GCGCGCATCG
	851	GTACGATTTT	GCTGCCGACT	TTGTCCAAAC	ACTCGGCAAA	CCAAGATACG
45	901	GAACAGTTTT	CCGCCCTGCT	CGACTGGGGT	TTGCGCNTGT	GCATGCTGCT
	951	GACGCTGCCG	GCGGCGGTCT	GAATGGCGGT	GTTGTCTGTT	CCGCTGGTGG
	1001	CAACCTTGTT	TATGTACCGA	GAATTCACGC	TGTTTGACGC	GCAGATGACG
	1051	CAACACGCGC	TGATTGCTTA	TTCTTTTCGT	TTAATCGGTT	TAATCATGAT
	1101	TAAAGTGTGG	GCGCCCGGCT	TTTATGCGCG	GCAAAACATC	AAAAAGCCCG
50	1151	TCAAAATCGC	CATCTTACAG	CTCATTTGCA	CGCAGTTGAT	GAACCTTGCC
	1201	TTTATCGGCC	CACTGAAACA	CGTCGGACTT	TCGCTTGCCA	TCGGTCTGGG
	1251	CGCGTGTATC	AATGCCGGAT	TGTTGTTTTC	CCTGTTGCGC	AGACACGGTA
	1301	TTTACCAACC	TGGCAAGGGT	TGGGACAGCT	TCTTTGGCAA	AATGCTGTCT
	1351	TCGCTCGCCG	TGATGGGAGG	CGGCTGTGAT	CCGCCCCAAA	TCTGGCTGCC
	1401	GTTTCGATCG	GCACACGCGG	GCGGAATGCA	AAAGGCGGCC	CGGCTCTTCA

1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GCGGGCTTTG
 1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA

This encodes a protein having amino acid sequence [SEQ ID 118] (SEQ ID NO: 118):

```

5      1  MNMLGALVKV  GSLTMVSRVL  GFVRDVIAR  AFGAGMATDA  FFVAFKLPNL
      51  LRRVFAEGAF  AQAFVPILAE  YKETRSKEAT  EAFIRHVAGM  LSFVLVIVTA
     101  LGILAAPWVI  YVSAPGFAKD  ADKFQLSIDL  LRITFPYILL  ISLSSFVGSV
     151  LNSYHKFSIP  AFTPTFLNVS  FIVFALFFVP  YFDPVVTALA  WAVFVGGILQ
     201  LGFQLPWLAK  LGFLKLPKLS  FKDAAVNRVM  KQMAPAILGV  SVAQISLVIN
    10  251  TIFASYLQSG  SVSWMYADR  MMELPGGVLG  AALGTILLPT  LSKHSANQDT
     301  EQFSALLDWG  LRXCMLLTLP  AAVGMAVLSF  PLVATLFMYR  EFTLFDAQMT
     351  QHALIAYSFG  LIGLIMIKVL  APGFYARQNI  KTPVKIAIFT  LICTQLMNLA
     401  FIGPLKHVGL  SLAIGLGACI  NAGLLFYLLR  RHGIYQPGKG  WAAFLAKMLL
     451  SLAVMGGGLY  AAQIWLPPDW  AHAGGMQKAA  RLFILIAVGG  GLYFASLAAL
    15  501  GFRPRHFKRV  ES*
  
```

ORF20a (SEQ ID NO: 118) and ORF20-1 (SEQ ID NO: 116) show 96.5% identity in 512 aa overlap:

```

20      10      20      30      40      50      60
    orf20a.pep  MNMLGALVKVGSLSLTMVSRVLGFVRDVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
    orf20-1      MNMLGALAKVGSLSLTMVSRVLGFVRDVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
                10      20      30      40      50      60

25      70      80      90      100     110     120
    orf20a.pep  AQAFVPILAELYKETRSKEATEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAKD
    orf20-1      AQAFVPILAELYKETRSKEAEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAQD
                70      80      90      100     110     120

30      130     140     150     160     170     180
    orf20a.pep  ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAFPTPTFLNVSFIVFALFFVP
    orf20-1      ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFPTPTFLNVSFIVFALFFVP
                130     140     150     160     170     180

35      190     200     210     220     230     240
    orf20a.pep  YFDPVVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
    orf20-1      YFDPVVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
                190     200     210     220     230     240

40      250     260     270     280     290     300
    orf20a.pep  SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLGAAALGTILLPTLSKHSANQDT
    orf20-1      SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAAALGTILLPTLSKHSANQDT
                250     260     270     280     290     300

45      310     320     330     340     350     360
    orf20a.pep  EQFSALLDWGLRXCMLLTLPAAVGMVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
    orf20-1      EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
                310     320     330     340     350     360
  
```


		370	380	390	400	410	420
	orf20a.pep	LIGLIMIKVLAPGFIYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKXVGLSLAIGLGACI					
5	orf20-1	LIGLIMIKVLAPGFIYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKXVGLSLAIGLGACI					
		370	380	390	400	410	420
	orf20a.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWLFPDWAHAGGMQKAA					
	orf20-1	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
10		430	440	450	460	470	480
	orf20a.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWLFPDWAHAGGMQKAA					
	orf20-1	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
		430	440	450	460	470	480
	orf20a.pep	RLFILIAVGGGLYFASLAALGFRPRHFKRVESX					
	orf20-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
15		490	500	510			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF20 (SEQ ID NO: 114) shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) (SEQ ID NO: 120) from *N. gonorrhoeae*:

20	orf20.pep	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20ng	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20.pep	AQAFVPILA EYK ETRSKEAXEAFIRHVAGMLS FVLVI V TALGILAAPWVIYVSAPSFAQD	120
	orf20ng	AQAFVPILA EYK ETRSKEATEAFIRHVAGMLS FVLIVV TALGILAAPWVIYVSAPGFTKD	120
25	orf20.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP	180
	orf20ng	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP	180
	orf20.pep	YFDPPVTAXAWAVFVG GILQLX FQLPWLAKL GFLKLPKLSFKDAAVNRVMQM A PAILGV	240
30	orf20ng	YFDPPVTALAWAVFVG GILQLG FQLPWLAKL GFLKLPKLNFKDAAVNRVMQM A PAILGV	240
	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYYADRM MELPSGVLGAALGTILLPTLSKHSANQDT	300
	orf20ng	SVAQISLVINTIFASYLQSGSVSWMYYADRM MELPGGVLGAALGTILLPTLSKHSANQDT	300
35	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG	360
	orf20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG	360
	orf20.pep	LIGLIMIKVLAPGFIYARQNIIXPVKIAIFTL ICXQLMNLXFXG PLXXIGLSLAIGLGACI	420
	orf20ng	LIGLIMIKVLASGFIYARQNIKTPVKIAIFTL ICTQLMNLAFIGPLKHAGLSLAIGLGACI	420
40	orf20.pep	NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSP	454
	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP	454

An ORF20ng nucleotide sequence [<SEQ ID 119>] (SEQ ID NO: 119) was predicted to encode a protein having amino acid sequence [<SEQ ID 120>] (SEQ ID NO: 120):

```

1  MNMLGALAKV  GSLTMVSRVL  GFVRD TVIAR  AFGAGMATDA  FFVAFKLPNL
5  51  LRRVFAEGAF  AQAFVPILAE  YKETRSKEAT  EAFIRHVAGM  LSFVLIVVTA
101 LGILAAPWVI  YVSAPGFTKD  ADKFQLSISL  LRITFPYILL  ISLSSFVGS I
151 LNSYHKFGIP  AFTPTFLNIS  FIVFALFFVP  YFDP PVTALA  WAVFVGGILO
201 LGFQLPWLAK  LGFLKLPKLN  FKDAAVNRVM  QMAPAILGV   SVAQISLVIN
251 TIFASYLQSG  SVSWMYADR   MMELPGGVLG  AALGTILLPT  LSKHSANQDT
301 EQFSALLDWG  LRLCMLLTLP  AAAGLAVLSF  PLVATLFMYR  EFTLFDAQMT
10 351 QHALIAYSFG  LIGLIMIKVL  ASGFYARQNI  KTPVKIAIFT  LICTQLMNL A
401 FIGPLKHAGL  SLAIGLGACI  NAGLLFFLFR  KHGIYRPGQG  LGQPSWRKCC
451 SRSP*

```

Further DNA sequence analysis revealed the following DNA sequence [<SEQ ID 121>] (SEQ ID NO: 121):

```

1  ATGAATATGC  TTGGAGCTTT  GGCAAAAGTC  GGCAGCCTGA  CGATGGTGTC
51  GCGCGTTTGTG  GGATTGTGTC  GCGATACGGT  CATTGCGCGG  GCATTCGGCG
101 CGGGTATGGC  GACGGATGCG  TTTTGTGTCG  CGTTCAAAC T  GCCCAACCTG
151 CTTCCGCCGCG  TGTTCGCGGA  GGGGGCGGTT  GCCCAAGCGT  TTGTGCCGAT
20 201 TTTGGCGGAA  TATAAGGAAA  CGCGTTCTAA  AGAGGCGA c gAGGCTTTTA
251 TCCGCCACGt  tgcgggAatg  CTGTCGTTTG  TGCTGATcgt  cGttacCGCG
301 CTGGGCATAC  TTGCCGCgcc  tGGGTGATT  TATGTTtccg  CgcccGGCTT
351 TACCAAAGAC  GCGGACAAGT  TCCAACTTTC  CATCAGCCTG  CTGCGGATTA
401 CGTTTCCTTA  TATATTATTG  ATTTCTTTGT  CTTCTTTTGT  CGGCTCGATA
25 451 CTAATTCCT  ACCATAAGTT  CGGCATTCCC  GCGTTTACGC  CCACGTTTTT
501 AAACATCTCT  TTTATCGTAT  TCGCACTGTT  TTTCTGCGG  TATTTCGATC
551 CGCCCGTTAC  CGCGCTGGCG  TGGGCGGTTT  TTGTCGGCGG  TATTTG CAG
601 CTCGGTTTCC  AACTGCCGTG  GCTGGCGAAA  CTGGGCTTTT  TGAAACTGCC
651 CAACTGAAT  TTCAAAGATG  CGGCGGTCAA  CCGCGTCATG  AACAGATGG
30 701 CGCCTGCGAT  TTTGGGCGTG  agcgTGGCGC  AAATTCTTT  GgttATCAAC
751 ACGATTTTCG  CGTCTTATCT  GCAATCGGGC  AGCGTTTCAT  GGATGTatta
801 cgCCGACCGC  ATGATGGAGc  tgcgccGGGG  CGTGCTGGGG  GCTGCACTCG
851 GTACAAATTT  GCTGCCGACT  TTGTCCAAAC  ACTCGGCAAA  CCAAGATACG
901 GAACAGTTT  CCGCCCTGCT  CGACTGGGGT  TTGCGCCTGT  GCATGCTGCT
35 951 GACGCTGCCG  GCGGCGGccg  GACTGGCGGT  ATTGTCGTTC  CCGCTGGTGG
1001 CGACGCTGTT  TATGTACCGA  GAATTCACGC  TGTTCGACGC  ACAAATGACG
1051 CAACACGCGC  TGATTGCCTA  TTCTTTCGGT  TTAATCGGTT  TAATTATGAT
1101 TAAAGTGTTG  GCATCCGGCT  TTTATGCGCG  GCAAAACATC  AAAACGCCCCG
1151 TCAAAATCGC  CATCTTCACG  CTCATCTGCA  CGCAGTTGAT  GAACCTCGCC
40 1201 TTTATCGGTC  CGTTGAAACA  CGCCGGGCTT  TCGCTCGCCA  TCGGCCTGGG
1251 CGCGTGATC  AACGCCGGAT  TGTGTTCTT  CCTGTTGCGC  AAACACGGTA
1301 TTTACCGGCC  cggcaggggt  tgggcggcgt  TCTTGGCGAA  AATGCTGCTC
1351 GCGCTCGCCG  TGATGTGCGG  CGGACTGTGG  GCGGCGCAGG  CTTGCCTGCC
45 1401 GTTCGAATGG  GCGCACGCCG  GCGGAATGCG  GAAAGCGGGG  CAGCTCTGCA
1451 TCCTGATTGC  CGTCGGCGGC  GGACTGTATT  TCGCATCTCT  GGCGGCTTTG
1501 GGCTTCCGTC  CGCGCCATTT  CAAACGCGTG  GAAAGCTGA

```

This encodes the following amino acid sequence [<SEQ ID 122; ORF20ng-1>] (SEQ ID NO: 122; ORF20ng-1):

```

50 1  MNMLGALAKV  GSLTMVSRVL  GFVRD TVIAR  AFGAGMATDA  FFVAFKLPNL
51 51  LRRVFAEGAF  AQAFVPILAE  YKETRSKEAT  EAFIRHVAGM  LSFVLIVVTA

```

101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGS
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPVPTALA WAVFVGGILQ
 201 LGFLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT
 5 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
 401 FIGPLKHAGL SLAIGLGACI NAGLLFLLR KHGIYRPGRG WAAFLAKMLL
 451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
 501 GFRPRHFKRV ES*

ORF20ng-1 (SEQ ID NO: 122) and ORF20-1 (SEQ ID NO: 116) show 95.7% identity in 512 aa overlap:

		10	20	30	40	50	60
15	orf20-1.pep	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF					
	orf20ng-1	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF					
		10	20	30	40	50	60
	orf20-1.pep	70	80	90	100	110	120
20	orf20ng-1	AQA FVPILA EYK ETR SKEA EAFIRHVAGMLS FVLVIVTALG ILAAPWVIYVSAPGFAQD					
		70	80	90	100	110	120
	orf20-1.pep	130	140	150	160	170	180
25	orf20ng-1	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIP AFTPTFLNVSFIVFALFFVP					
		130	140	150	160	170	180
	orf20-1.pep	190	200	210	220	230	240
30	orf20ng-1	YFDPVPTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV					
		190	200	210	220	230	240
	orf20-1.pep	250	260	270	280	290	300
35	orf20ng-1	SVAQVSLVINTIFASYLQSGSVSWMYADRM MELPSGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
	orf20-1.pep	310	320	330	340	350	360
40	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQH ALIAYSFG					
		310	320	330	340	350	360
	orf20-1.pep	370	380	390	400	410	420
45	orf20ng-1	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTL ICTQLMNLAFIGPLKHVGLSLAIGLGACI					
		370	380	390	400	410	420
	orf20-1.pep	430	440	450	460	470	480
50	orf20ng-1	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					

```

orf20ng-1      NAGLLFFLLRKHGIYRPGRGWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRKAG
                430      440      450      460      470      480

                490      500      510
orf20-1.pep    QLCILIAVGGGLYFASLAALGFRPRHFKRVENX
5              |||||
orf20ng-1      QLCILIAVGGGLYFASLAALGFRPRHFKRVESX
                490      500      510

```

In addition, ORF20ng-1 (SEQ ID NO: 122) shows significant homology with a virulence factor (SEQ ID NO: 1122) of *S.typhimurium*:

```

sp|P37169|MVIN_SALTY VIRULENCE FACTOR MVIN pir||S40271 mviN protein - Salmonella
typhimurium gi|438252 (Z26133) mviB gene product [Salmonella typhimurium]
gnl|PID|d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524
Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
Identities = 309/467 (66%), Positives = 368/467 (78%)

```

```

Query:      1 MNMLGALAKVGSMTVSRVLGFVRDVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60
             MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF
Sbjct:     14 MNLLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73

```

```

Query:     61 AQAFVPILAIEYKETSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD 120
             +AQAFVPILAIEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF
Sbjct:     74 SQAFVPILAIEYKSKQGEETRIFVAYVSGLLTLALAVVTAGMLAAPWVIMVTAPGFADT 133

```

```

Query:    121 ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTFLNISFIVFALFFVP 180
             ADKF L+ LLRITFPYILLISL+S VG+ILN++++F IPAF PTFLNIS I FALF P
Sbjct:    134 ADKFALTTQLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193

```

```

Query:    181 YFDPVPTALAWAVFVGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMQMAPAILGV 240
             YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+QQM PAILGV
Sbjct:    194 YFNPPVLALAWAVTVGGVLQVLYQLPYLKKIGMLVLPRLNFRDGTAMRVVQMGPAILGV 253

```

```

Query:    241 SVAQISLVINTIFASYLQSGSVSWMYYADRMELRRGVLGAAALGTILLPTLSKHSANQDT 300
             SV+QISL+INTIFAS+L SGSVSWMYYADR+ME GVLG ALGTILLP+LSK A+ +
Sbjct:    254 SVSQISLIINTIFASFLASGSVSWMYYADRLMEFPSPGVLGVALGTILLPSLSKSFASGNH 313

```

```

Query:    301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG 360
             +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G
Sbjct:    314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYQKFTAFDAAMTQRALIAYSVG 373

```

```

Query:    361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420
             LIGLI++KVLG GFY+RQ+IKTPVKIAI TLI TOLMNLAFIGPLKHAGLSL+IGL AC+
Sbjct:    374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

```

```

Query:    421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACLP 467
             NA LL++ LRK I+ P GW VM L+ +P
Sbjct:    434 NASLLYWQLRKQNIPTPQPGWWMFLMRLIISVLVMAAVLFGVLHIMP 480

```

```

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
Identities = 14/41 (34%), Positives = 23/41 (56%)

```

```

Query:    469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509
             EW+ + + +L ++ G YFA+LA LGF+ + F R
Sbjct:    481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFVKVKEFVR 521

```

Based on this analysis, including the homology with a virulence factor (SEQ ID NO: 1122) from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 15

- 5 The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 123] (SEQ ID NO: 123):

```

1  atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
51  GCAAGCCGTT tACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGGCG
101 AAGAAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
10  151 GTcAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGGCGAAA
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAArGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAAC TTAA GCGGCGAAGA
351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
15  401 GTCCGTTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAAATGCGA tGGACACCAA TCCG..

```

This corresponds to the amino acid sequence [SEQ ID 124; ORF22] (SEQ ID NO: 124; ORF22):

```

20 1 MIKIKKGLNL PIAGRPEQAV YDGPAITEVA LLGEEYAGMR PSMKVKEGDA
51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEXNDEI
101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151 VNAMD TNP..

```

- 25 Further work revealed the complete nucleotide sequence [SEQ ID 125] (SEQ ID NO: 125):

```

1  ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
51  GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
101 AAGAAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
30  151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTAC CGTGGCGAAA
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAAC TTAA GCGGCGAAGA
351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
35  401 GTCCGTTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATTAT
501 CAAAGAAAGCC GCCGAGGATT TCAAACGCGG CCTGTGTGTA TTGAGCGGTT
551 TGACCGAACG CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
651 TGCCGGTTTG AGTGGCACGC ACATTCATTT CATCGAGCCG GTCGGCGCGA
40  701 ATAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
801 TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTGG GGTGCGAAAG
851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
45  951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
1001 AGCTGTTTCG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT

```

1051 ACAACCCTCG GCCATTTCTT GAAAAACAAA CTCTTCAAGT TCAACACAGC
 1101 CGTCAACGGC GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTGCGCGA TTTAATCGTC
 1201 GGCGATACCG ACAGCGCGCA GGCATTGGGT TGCTTGGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence [<SEQ ID 126; ORF22-1>] (SEQ ID NO: 126; ORF22-1):

1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMD TNPLA ADPTV IIEKA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFI EP VGANKTVWTI NYQDVITIGR
 15 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI
 301 SGSVLNGAIT QGAHDYLG RY HNQISVIEEG RSKELFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

Further work identified the corresponding gene in strain A of *N.meningitidis* [<SEQ ID 127>] (SEQ ID NO: 127):

1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
 25 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
 201 GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA
 251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACCTTA GCGGCGANGA
 351 ANTNNGNNGC AATCTGATCC AATCCGGTTT GTGGA CTGCGTANCC
 30 401 GTCCGTTTCAG CAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTTGTGAT
 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
 551 TGACCGAGCG TAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCGC
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTGCGCG GCCCGCATCC
 35 651 GGCCGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGGTGCAA
 701 ACAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
 801 TTCTCAAGTC AACAAACAC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
 851 TATCGCAAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
 40 901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAATACTC CATCACGCGT
 1051 ACGACCCTCG GCCATTTCTT GAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGT GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 45 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGC CGA TTTAATCGTC
 1201 GGCGATACCG ACAGCGCGCA AGCATTGGGT TGCTTGGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
 1301 CGCTGTTGCG TAAGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence [<SEQ ID 128; ORF22a>] (SEQ ID NO: 128; ORF22a):

1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PXMKVKEGDA
 51 VKKGQVLFED KKXPGVVFTA PVSCKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGXEXXX NLIQSLWTA LRXPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPVVVIKEA XXDFRXXLV LSRLTERKIH VCKAAGADV
 201 SENAANIETH EFGGPHAGL SGTHHFIEP VGANKTVWTI NYQDVIAIGR
 251 LFATGRLNTE RVIALGGSQV NKPRLRLTVL GAKVSQITAG ELVDADNRVI
 301 SGSVLNGAIT QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLRLRLDIL
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*

The originally-identified partial strain B sequence (ORF22) (SEQ ID NO: 124) shows 94.2% identity over a 158aa overlap with ORF22a (SEQ ID NO: 128):

		10	20	30	40	50	60
orf22.pep		MIKIKKGLNLPIAGRPEQAVYDGP	ITEVALLGEEYAGMRPSMKVKEGDAV	KKKGQVLFED			
orf22a		MIKIKKGLNLPIAGRPEQVIYDGP	VITEVALLGEEYAGMRPXMKVKEGDAV	KKKGQVLFED			
		10	20	30	40	50	60
orf22.pep		KKNPGVVFTAPASGKIAAIHRGEKRVLSV	VI	AVEGNDEIEFERYAPEALANLSGEEVRR			
orf22a		KKXPGVVFTAPVSGKIAAIHRGEKRVLSV	VI	AVEGNDEIEFERYAPEALANLSGXEXXX			
		70	80	90	100	110	120
orf22.pep		NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP					
orf22a		NLIQSLWTALRXPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV					
		130	140	150	160	170	180

The complete strain B sequence (ORF22-1) (SEQ ID NO: 126) and ORF22a (SEQ ID NO: 128) show 94.9% identity in 447 aa overlap:

		10	20	30	40	50	60
orf22a.pep		MIKIKKGLNLPIAGRPEQVIYDGP	VITEVALLGEEYAGMRPXMKVKEGDAV	KKKGQVLFED			
orf22-1		MIKIKKGLNLPIAGRPEQAVYDGP	ITEVALLGEEYAGMRPSMKVKEGDAV	KKKGQVLFED			
		10	20	30	40	50	60
orf22a.pep		KKXPGVVFTAPVSGKIAAIHRGEKRVLSV	VI	AVEGNDEIEFERYAPEALANLSGXEXXX			
orf22-1		KKNPGVVFTAPASGKIAAIHRGEKRVLSV	VI	AVEGNDEIEFERYAPEALANLSGEEVRR			
		70	80	90	100	110	120
orf22a.pep		NLIQSLWTALRXPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV					
orf22-1		NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV					
		130	140	150	160	170	180
orf22a.pep		LSRLTERKIHVCKAAGADVPSENAANIETH	EFGGPHAGLSGTHHFIEPVGANKTVWTI				
orf22-1		LSRLTERKIHVCKAAGADVPSENAANIETH	EFGGPHAGLSGTHHFIEPVGANKTVWTI				
		190	200	210	220	230	240

-160-

		190	200	210	220	230	240
		250	260	270	280	290	300
	orf22a.pep	NYQDVIAIGRLFATGRLNTERVIALGGSQVKNPRLRLRTVLGAKVSQITAGELVDADNRVI					
5	orf22-1	NYQDVITIGRLFATGRLNTERVIALGGSQVKNPRLRLRTVLGAKVSQITAGELVDTDNRVI					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf22a.pep	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFLKNK					
10	orf22-1	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFLKNK					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf22a.pep	LFKFTTAVNGGDRAMVPIGTYERVMPDLIPTLLLRDLIVGDTDSAQALGCLELDEEDLA					
15	orf22-1	LFKFNTAVNGGDRAMVPIGTYERVMPDLIPTLLLRDLIVGDTDSAQALGCLELDEEDLA					
		370	380	390	400	410	420
		430	440				
	orf22a.pep	LCSFVCPGKYEXGPLLRRKVLETXEKEGX					
20	orf22-1	LCSFVCPGKYEYGPLLRRKVLETIEKEGX					
		430	440				

Further work identified a partial gene sequence [[SEQ ID 129](#)] ([SEQ ID NO: 129](#)) from *N.gonorrhoeae*, which encodes the following amino acid sequence [[SEQ ID 130; ORF22ng](#)] ([SEQ ID NO: 130; ORF22ng](#)):

25	1	MIKIKKGLNL	PIAGRPEQVI	YDGPATEVA	LLGEEYVGM	PSMKIKEGEA
	51	VKKGQVLFED	KKNPGVVFTA	PASGKIAAIH	RGEKRVLSV	VIAVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF
	151	VNAMDTNPLA	ADPTVIIKEA	AEDFKRGLLV	LSRLTERKIH	VCKAAGADV
	201	SENAANIETH	EFGGPHAGL	SGTHIHFI	VGANKTVWTI	NYQDVIAIGR
30	251	LFVTGRLNTE	RVVALGGLQV	NKPRLRLTVL	GAKVSQTLTAG	ELVDADNRVI
	301	SGSVLNGAIA	QGAHDYLG	RYHN*		

Further work identified complete gonococcal gene [[SEQ ID 131](#)] ([SEQ ID NO: 131](#)):

	1	ATGATTAAAA	TCAAAAAAGG	TCTAAATCTG	CCCATCGCGG	GCAGACCGGA
35	51	GCAAGTCATT	TATGACGGCC	CGGCCATTAC	CGAAGTCGCG	TTGCTTGCGG
	101	AAGAATATGT	CGGCATGCGC	CCCTCGATGA	AAATCAAGGA	AGGTGAAGCC
	151	GTCAAAAAAG	GCCAAGTGCT	GTTTGAAGAC	AAAAAGAATC	CGGGCGTAGT
	201	ATTTACTGCG	CCGGCTTCAG	GCAAAATCGC	CGCTATTAC	CGTGGCGAAA
	251	AGCGCGTACT	TCAGTCAGTC	GTGATTGCCG	TTGAAGGCAA	CGACGAAATC
40	301	GAGTTCGAAC	GCTACGTACC	TGAAGCGCTG	GCAAAATTGA	GCAGCGAAAA
	351	AGTGCGCCGC	AACCTGATTC	AATCAGGCTT	ATGGACTGCG	CTTCGCACCC
	401	GTCCGTTTCA	CAAAATCCCT	GCCGTAGATG	CCGAGCCGTT	CGCCATCTTC
	451	GTCAATGCGA	TGGACACCAA	TCCGCTGGCT	GCCGACCCTA	CGGTCATCAT
	501	CAAAGAGGCC	GCCGAAGACT	TCAAACGCGG	CCTGTTGGTA	TTGAGCCGCC
45	551	TGACCGAACG	TAAATCCAT	GTGTGTAAAG	CAGCAGGCGC	AGACGTGCCG
	601	TCTGAAAATG	CTGCCAATAT	CGAAACACAT	GAATTGGCG	GCCCGCATCC
	651	TGCCGCTTG	AGTGGCACGC	ACATTCAATT	CATCGAGCCA	GTGCGCGCGA
	701	ATAAAACCGT	GTGGACCATC	AATTATCAAG	ACGTGATTGC	TATCGGACGT
	751	TTGTTCTGTA	CAGGCCGTCT	GAATACCGAG	CGCGTGTTTG	CCTTGGGCGG
50	801	CCTGCAAGTC	AACAAACCGC	GCCTCTTGCG	TACCGTTTTG	GGTGCGAAGG

-162-

		70	80	90	100	110	120
		130	140	150	160	170	180
	orf22-1.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV					
5	orf22ng-1	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf22-1.pep	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTVWTI					
10	orf22ng-1	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTVWTI					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf22-1.pep	NYQDVITIGRLFATGRLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDTDNRVI					
15	orf22ng-1	NYQDVIAIGRLFVTGRLNTERVVALGGLQVKNPRLRLTVLGAKVSQITAGELVDADNRVI					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf22-1.pep	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQDPKYSITRTLGHFLKNK					
20	orf22ng-1	SGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQDPKYSITRTLGHFLKNK					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf22-1.pep	LFKFNTAVNGGDRAMVPIGTYERVMLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA					
25	orf22ng-1	LFKFNTAVNGGDRAMVPIGTYERVMLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA					
		370	380	390	400	410	420
		430	440				
	orf22-1.pep	LCSFVCPGKYEYGPLLRRKVLETIEKEGX					
30	orf22ng-1	LCSFVCPGKYEYGPLLRRKVLETIEKEGX					
		430	440				

Computer analysis of these sequences gave the following results:

35 Homology with 48kDa outer membrane protein of Actinobacillus pleuropneumoniae (accession number U24492) (SEQ ID NO: 1123).

ORF22 (SEQ ID NO: 124) and this 48kDa protein (SEQ ID NO: 1123) show 72% aa identity in 158aa overlap:

40	orf22	1	MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED	60
	48kDa	1	MI IKKGL+LPIAG P Q +++G + EVA+LGEEY GMRPSMKV+EGD VKKGQVLFED	60
	orf22	61	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIIVAEVXNDEIEFERYAPEALANLSGEEVRR	120
	48kDa	61	KKNPGVVFTAPASG + I+RGEKRVLSVVI VE +++I F RY LA+LS E+V++	120

orf22 121 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158
 NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNP
 48kDa 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNP 158

ORF22a (SEQ ID NO: 128) also shows homology to the 48kDa *Actinobacillus pleuropneumoniae*
 5 protein (SEQ ID NO: 1123):

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]
 Length = 449

Score = 530 bits (1351), Expect = e-150
 Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

10 Query: 1 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED 60
 MI IKKGL+LPIAG P QVI++G + EVA+LGEEY GMRP MKV+EGD VKKGQVLFED
 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

Query: 61 KKPXPGVVFTAPVSGKIAAIHRGEKRVLSVVIIVEGNDEIEFERYAPEALANLSGXEXXX 120
 KK PGVVFTAP SG + I+RGEKRVLSVVI VEG+++I F RY LA+LS +
 15 Sbjct: 61 KKNPGVVFTAPASGTVVITINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVQKQ 120

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV 180
 NLI+SGLWTA R RPFSK+PA+DA P +IFVNAMDTNPLAADP VV+KE DF+ V
 Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGHPAGLSGTHIHFIIEPVGANKTV 237
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHF++PVGA K V
 20 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240

Query: 238 WTINYQDVIAIGRLFATGRLINTERVIALGGSQVKNPRLRLRTLVLGAKVSQITAGELVDADN 297
 W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQ+TA EL +N
 Sbjct: 241 WHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVRLGANLSQLTANELNAGEN 300

Query: 298 RVISGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFL 357
 RVISGSVL+GA G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF
 25 Sbjct: 301 RVISGSVLGATAAGPVDYLGRYALQSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KNKLFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXXXXVGDTSQAQXXXXXXXXXX 417
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ
 30 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTSQAQNLGCLELDEE 419

Query: 418 XXXXSFVCPGKYEXGPLLRKVLETXEKEG 447
 ++VCPGK GP+LR LE EKEG

ORF22ng-1 (SEQ ID NO: 132) also shows homology with the OMP (SEQ ID NO: 1123) from
A. pleuropneumoniae:

35 gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus
 pleuropneumoniae] Length = 449
 Score = 555 bits (1414), Expect = e-157
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

40 Query: 27 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFED 86
 MI IKKGL+LPIAG P QVI++G + EVA+LGEEYVGMRPSMK++EG+ VKKGQVLFED
 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 146
 KKNPGVVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++
 Sbjct: 61 KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 147 NLIQSGLWLTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 206
 NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNPLAADP V++KE DFK GL V
 Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 263
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIHF++PVGA K V
 Sbjct: 181 LTRLFNGQKPVYLCADSDNIPLSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTGRINTERVVALGGLQVKNPRLRLTVLGAKVSQLTAGELVDADN 323
 W +NYQDVIAIG+LF TG L T+R+++L G QV PRL+RT LGA +SQLTA EL +N
 Sbjct: 241 WHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

Query: 324 RVISGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFL 383
 RVISGSVL+GA A G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF
 Sbjct: 301 RVISGSVLGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KKNLKFKFTTAVNGGDRAMVPIGTIERVMXXXXXXXXXXXXXXXXXVGDTSQAQXXXXXXXXX 443
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ
 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIIPTLLLRDLAAGDTSQAQNLGCLELDEE 419

Query: 444 XXXXXSFVCPGKYEYGPLLRKVLETIEKEG 473
 ++VCPGK YGP+LR LE IEKEG
 Sbjct: 420 DLALCTYVCPGKNNGPMLRAALEKIEKEG 449

Based on this analysis, including the homology with the outer membrane protein (SEQ ID NO: 1123) of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (SEQ ID NO: 126) (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 (SEQ ID NO: 126) is a surface-exposed protein, and that it is a useful immunogen.

Example 16

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 133] (SEQ ID NO: 133):

```

      1  ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
    51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTATTTTGT
  101  TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
  151  TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
    5   201  ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
  251  TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
  301  ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
  351  TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
  401  GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
    10  451  ACTCTGGsGC TTTmTTGsw CAkCATCTTT TTTGCCGCAC AGTTTGTCCG
  501  ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
  551  CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTGTGT TATCGGTTTT
  601  ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
  651  ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
    15  701  ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
  751  AATATTATTA CGCCGATGAT GAGTTATTTT GGGCTGATTA TGGCGACGGT
  801  GrkCmmTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
  851  TGCCGTATTC CGCTTCTTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
  901  TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
    20  951  TCCGCACCT TAA

```

This corresponds to the amino acid sequence [<SEQ ID 134; ORF12>] (SEQ ID NO: 134; ORF12):

```

    25  1  ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
      51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
  101  TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
  151  TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIFG
  201  ILICAFINLM IGSASAQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
  251  NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
    30  301  WVFVLGLPVG PGAPTFYPAP *

```

Further sequence analysis revealed the complete DNA sequence [<SEQ ID 135>] (SEQ ID NO: 135) to be:

```

    35  1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTT ATTATTTTCA
  101  TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
  151  GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
  201  GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
  251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTG
    40  301  GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
  351  ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAATC ACTACTTTTA
  401  TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
  451  GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
  501  TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
    45  551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
  601  CAACAGGCGG CGCAAATCAT CCATCCGAC TACGTCGTAG GCCCTGAAGC
  651  CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
  701  ATTTTGTTAC TGAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
  751  GATTGTGCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCAGGCC
    50  801  TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTT GTTGCCTTAT
  851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
  901  CCTGAAACAG GATTGGTTTC CGGTCGCCG TTTTAAAAAT CGATTGTTGT
  951  TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATTGTTTAT GGCCGGGTAA
  1001  CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
    55  1051  ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT

```

5	1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	GCCGTTAAAG
	1151	GGGCGACGTT	CTTAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC
	1201	GGTTTTATTT	TAATTTGTGC	TTTTATCAAT	CTGATGATAG	GCTCCGCCCTC
	1251	CGCGCAATGG	GCGGTAAC TG	CGCCGATTTT	CGTCCCTATG	CTGATGTTGG
	1301	CCGGCTACGC	GCCC GAAGTC	ATTCAAGCCG	CTTACCGCAT	CGGTGATTCC
	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCCGGC	TGATTATGGC
	1401	GACGGTGATC	AAATACAAAA	AAGATGCGGG	CGTGGGTACG	CTGATTCTTA
	1451	TGATGTTGCC	GTATTCCGCT	TTCTTCTTGA	TTGCGTGGAT	TGCCTTATTC
10	1501	TGCATTTGGG	TATTTGTTT	GGGCCTGCC	GTGCGTCCCG	GCGCGCCAC
	1551	ATTCTATCCC	GCACCTTAA			

This corresponds to the amino acid sequence [[<SEQ ID 136; ORF12-1>](#)] ([SEQ ID NO: 136; ORF12-1](#)):

15	1	MSQTD	TQRDG	RFLRT	VEWLG	NMLPH	PVTLF	IIFIV	LLLIA	SAVGAY	FGLS
	51	VPDPR	PVGAK	GRADD	GLIYI	VSLLN	ADGFI	KILTH	TVKNF	TGFAPL	GTVL
	101	VSLLG	VGIAE	KSGLI	SALMR	LLLT	KSPRK	LT	FMV	VFTGI	LSNTASE
	151	VVLIPL	SAII	FHSLG	RHPLA	GLAA	FAGVS	GGY	SAN	FLG	TIDPLLAGIT
	201	QQA	AQIIHPD	YVVG	PEANWF	FMV	ASTFVIA	LIGY	FVTEKI	VEPQLG	PHYQS
20	251	DLS	QEEKDIR	HSNEI	TPLEY	KGLIW	AGVVF	VALS	ALLAWS	IVPADG	GILRH
	301	PETGL	VSGSP	FLKSI	VVFIF	LLFAL	PGIVY	GRVTR	SLRGE	QEVV	NMAES
	351	MSTL	GLYLVI	IFFAA	QVFAF	FNW	TNIGQYI	AVK	GATFLKE	VGLGGS	VLFPI
	401	G	FILICAFIN	LMIGS	SASAQW	AVT	APIFVPM	LM	L	AGYAPEV	IQAAYRIGDS
	451	VTNI	IITPMMS	YFGLI	MATVI	KYK	KDAGVGT	LIS	MMLPYSA	FFLIAW	IALF
	501	CIWV	FVLGLP	VGPG	APTFFYP	AP*					

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 (SEQ ID NO: 134) shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) (SEQ ID NO: 138) from strain A of *N. meningitidis*:

[illegible]

-167-

		160	170	180	190	200	210
	orf12.pep	TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM					
	orf12a	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM					
5		360	370	380	390	400	410
		220	230	240	250	260	270
	orf12.pep	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIIITPMMSYFGLIMATVXXY					
	orf12a	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIIITPMMSYFGLIMATVIKY					
10		420	430	440	450	460	470
		280	290	300	310	320	
	orf12.pep	KKDAGVGTLIXMMLPYSAFFLIAWIALFCIWVFLGGLPVGPGAPTFFYPAPX					
	orf12a	KKDAGVGTLISMMMLPYSAFFLIAWIALFCIWVFLGGLPVGPGAPTFFYPAPX					
15		480	490	500	510	520	

The complete length ORF12a nucleotide sequence [[<SEQ ID 137>](#)] ([SEQ ID NO: 137](#)) is:

	1	ATGAGTCAAA	CCGATACGCA	ACGGGACGGA	CGATTTTAC	GCACAGTCGA
	51	ATGGCTGGGC	AATATGTTGC	CGCACCCGGT	TACGCTTTT	ATTATTTTCA
20	101	TTGTGTTATT	GCTGATTGCC	TCTGCCGCCG	GTGCGTATTT	CGGACTATCC
	151	GTCCCCGATC	CGCGCCCTGT	TGGTGCAGAA	GGACGTGCCG	ATGACGGTTT
	201	GATTCACGTT	GTCAGCCTGC	TCGATGCTGA	CGGTTTGATC	AAAATCTGA
	251	CGCATACCGT	TAAAAATTTC	ACCGGTTTCG	CGCCGTTGGG	AACGGTGTTG
	301	GTTTCTTTAT	TGGGCGTGGG	GATTGCGGAA	AAATCGGGCT	TGATTCCGC
25	351	ATTAATGCGC	TTATTGCTCA	CAAAATCTCC	ACGCAAACTC	ACTACTTTTA
	401	TGGTTGTTT	TACAGGGATT	TTATCTAATA	CCGCTTCTGA	ATTGGGCTAT
	451	GTCGTCCTAA	TCCCTTTGTC	CGCCATCATC	TTTCATTCCC	TCGGCCGCCA
	501	TCCGCTTGCC	GGTCTGGCTG	CGGCTTTCGC	CGGCGTTTCG	GGCGGTTATT
	551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTGGC	AGGCATCACC
30	601	CAACAGGCGG	CGCAAATCAT	CCATCCCGAC	TACGTCGTAG	GCCCTGAAGC
	651	CAACTGGTTT	TTTATGGTAG	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
	701	ATTTTGTAC	TGAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA
	751	GATTTGTCAC	AAGAAGAAAA	AGACATTCCA	CATTCCAATG	AAATCACGCC
	801	TTTGGAATAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTTT	GTTGCCTTAT
35	851	CCGCCCTATT	GGCTTGGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCGTCAT
	901	CCTGAAACAG	GATTGGTTTC	CGGTTTCGCG	TTTTTAAAT	CAATTGTTGT
	951	TTTTATTTTC	TTGTTGTTTG	CACTGCCGGG	CATTGTTTAT	GGCCGGGTAA
	1001	CCCGAAGTTT	GCGCGGCGAA	CAGGAAGTCG	TTAATGCGAT	GGCCGAATCG
	1051	ATGAGTACTC	TGGGGCTTTA	TTTGGTCATC	ATCTTTTTTG	CCGCACAGTT
40	1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	GCCGTAAAG
	1151	GGGCGACGTT	CTTAAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC
	1201	GGTTTTATTT	TAATTTGTGC	TTTTATCAAT	CTGATGATAG	GCTCCGCCTC
	1251	CGCGCAATGG	GCGGTAACGT	CGCCGATTTT	CGTCCCTATG	CTGATGTTGG
	1301	CCGGCTACGC	GCCCGAAGTC	ATTCAAGCCG	CTTACCGCAT	CGGTGATTCC
45	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCCGGC	TGATTATGGC
	1401	GACGGTGATC	AAATACAAAA	AAGATGCGGG	CGTGGGTACG	CTGATTTCTA
	1451	TGATGTTGCC	GTATTCCGCT	TTCTTCTTGA	TTGCGTGGAT	TGCCTTATTC
	1501	TGCATTTGGG	TATTTGTTTT	GGGCTGCCCG	GTCGGTCCCC	GCGCGCCAC
50	1551	ATTCTATCCC	GCACCTTAA			

This encodes a protein having amino acid sequence [[<SEQ ID 138>](#)] ([SEQ ID NO: 138](#)):

	1	MSQTDTRQD	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAAGAYFGLS
	51	VPDPRPVGAK	GRADDGLIHV	VSLLDADGLI	KILTHTVKNF	TGFAPLGTVL
	101	VSLLGVGIAE	KSLGISALMR	LLLTKSPRKL	TTFMVVFTGI	LSNTASELGY

151 VVLIPLSAII FHSLGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
 201 QQAAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
 251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
 301 PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
 351 MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLFI
 401 GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
 451 VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
 501 CIWVFLGLP VGPGAPTFYP AP*

ORF12a (SEQ ID NO: 138) and ORF12-1 (SEQ ID NO: 136) show 99.0% identity in 522 aa

overlap:

		10	20	30	40	50	60
orf12a.pep		MSQTD TQRDGRFLRTVEWLGNM LPHVTLFIIFIVLLLIASAAGAYFGLSVDPDRPVGAK					
orf12-1		MSQTD TQRDGRFLRTVEWLGNM LPHVTLFIIFIVLLLIASAVGAYFGLSVDPDRPVGAK					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf12a.pep		GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSL LGVGIAEKSGLISALMR					
orf12-1		GRADDGLIYIVSLLNADGFIKILTHTVKNFTGFAPLGTVLVSL LGVGIAEKSGLISALMR					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf12a.pep		LLLT KSPRKLTTFMVVTGILSNTASELGYVVL IPLSAIIFHSLGRHPLAGLAAAFAGVS					
orf12-1		LLLT KSPRKLTTFMVVTGILSNTASELGYVVL IPLSAIIFHSLGRHPLAGLAAAFAGVS					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf12a.pep		GGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
orf12-1		GGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf12a.pep		VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
orf12-1		VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
		250	260	270	280	290	300
		310	320	330	340	350	360
orf12a.pep		PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI					
orf12-1		PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI					
		310	320	330	340	350	360
		370	380	390	400	410	420
orf12a.pep		IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLMIGSASAQW					
orf12-1		IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLMIGSASAQW					
		370	380	390	400	410	420
		430	440	450	460	470	480
orf12a.pep		AVTAPIFVPM LMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					


```

orf12-1      AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIIITPMMSYFGLIMATVIKYKKGAGVGT
              430          440          450          460          470          480

              490          500          510          520
orf12a.pep   LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFYPAPX
orf12-1      LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFYPAPX
              490          500          510          520

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 (SEQ ID NO: 134) shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) (SEQ ID NO: 140) from *N. gonorrhoeae*:

orf12.pep	AXXI IHPXXVVGPEANWFFMVASTFVIALI	30
orf12ng	AAAFAGVSGGYSANLFLGTIDPLLAGITQQAQI IHPDYVVGPEANWFFMAASTFVIALI	232
orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	90
orf12ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	292
orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFI FLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	150
orf12ng	PADGILRHPETGLVAGSPFLKSIVVFI FLLFALPGIVYGRITRSLRGEREVVNAMAESMS	352
orf12.pep	TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM	210
orf12ng	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGAVFLKKFRLGGSVLFIFIGFILICAFINLM	412
orf12.pep	IGSASAQWAVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIIITPMMSYFGLIMATVXXY	270
orf12ng	IGSASAQWAVTAPIFVPMLMLAGNAPQVIQAAYRIGDSVTNIIITPMMSYFGLIMATVIKY	472
orf12.pep	KKDAGVGTLIXMMLPYSAFFLI AWIALFCIWWFVLGLPVGPGAPTFTYPAP	320
orf12ng	KKDAGVGTLISMMLPYSAFFLI AWIALFCIWWFVLGLPVGPGTPTFTYPVP	522

The complete length ORF12ng nucleotide sequence [<SEQ ID 139>] (SEQ ID NO: 139) is:

1	ATGAGTCAAA	CCGACGCGCG	TCGTAGCGGA	CGATTTTAC	GCACAGTCGA
51	ATGGCTGGGC	AATATGTTGC	CGCACC CGGT	TACGCTTTT	ATTATTTTCA
101	TTGTGTTATT	GCTGATTGcc	tctgCGTTCG	GTGCGTATTT	CGGACTATCC
151	GTCCCCGATC	CGCGTCCTGT	TGGGGCGAAA	GGACGTGCGG	ATGACGGTTT
201	GATTACAGTT	GTCAAGCTGT	TCGATGCCGA	CGGTTTGATC	AAAACTCTGA
251	CGCATACCGT	TAAAAATTTC	ACCGGTTTCG	CGCCGTTGGG	AACGGTGTGG
301	TTTTCTTTAT	TGGCGCTGGG	GATTGCGGAA	AAATCGGGCT	TGATTTCCGC
351	ATTAATGCGC	TTATTGCTCA	CAAAATCCCC	ACGCAAACTC	ACTACTTTTA
401	TGGTTGTTTT	TACAGGGATT	TTATCCAATA	CGGCTTCTGA	ATTGGGCTAT
451	GTCGTCCTAA	TCCCTTTGTC	CGCCGTCATC	TTTCATTTCG	TCGGCCGCCA
501	TCCGCTTGCC	GGTTTGGCTG	CGGCTTTCGC	CGGCGTTTCG	GGCGGTTATT
551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTTGCG	AGGCATCACC
601	CAACAGGCGC	CGCAAAATCA	CCATCCCCGAC	TACGTCGTAG	GCCCTGAAGC
651	CAACTGGTTT	TTTATGCGAC	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
701	ATTTTGTATC	TGAAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA

-170-

5 751 GATTGTGCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
 801 TTTGGAATAT AAAGGATTAA TTTGGGCAGG CGTGGTGTTT GTTGCCTTAT
 851 CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
 901 CCTGAAACAG GATTGGTTGC CGGTTCGCCG TTTTAAAAAT CGATTGTTGT
 951 TTTTATTTTC TTGTTGTTTG CGCTGCCGGG CATTGTTTAT GGCCGGATAA
 1001 CCCGAAGTTT GCGCGGCGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
 1051 ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
 1101 TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
 1151 GGGCGGTGTT CTTAAAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTTTATC
 1201 GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCTC
 1251 CGCGCAATGG GCGGTAAGT GCGCGATTTT CGTCCCTATG CTGATGTTGG
 1301 CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
 1351 GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
 1401 GACGGTAATC AAATACAAAA AAGATGCGGG CGTAGGCACG CTGATTTCTA
 1451 TGATGTTGCC GTATTCGCTT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
 1501 TGCATTTGGG TATTTGTTTT GGGTCTGCCC GTCGGTCCCC GCACACCCAC
 1551 ATTCTATCCG GTGCCTTAA

This encodes a protein having amino acid sequence [<SEQ ID 140>] (SEQ ID NO: 140):

20 1 MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
 51 VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILHTVKNF TGFAPLGTVL
 101 VSLLGVGIAE KSGLISALMR LLLTKSPRKL TTFMVVFTGI LSNASELGY
 151 VVLIPLSAVI FHSLSGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
 201 QQAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
 25 251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
 301 PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNAMAES
 351 MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGAVFLKK FRLGGSVLFI
 401 GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
 451 VTNIIITPMS YFGLIMATVI KYKDGAGVGT LISMMLPYSA FFLIAWIALF
 30 501 CIWVFVLGLP VPGTPTFYF VP*

ORF12ng (SEQ ID NO: 140) shows 97.1% identity in 522 aa overlap with ORF12-1 (SEQ ID NO: 136):

35	orf12-1.pep	MSQTD	10	20	30	40	50	60
	orf12ng	MSQTD	10	20	30	40	50	60
40	orf12-1.pep	GRADDGLI	70	80	90	100	110	120
	orf12ng	GRADDGLI	70	80	90	100	110	120
45	orf12-1.pep	LLLTKSPRKL	130	140	150	160	170	180
	orf12ng	LLLTKSPRKL	130	140	150	160	170	180
50	orf12-1.pep	GGYSANLFLG	190	200	210	220	230	240
	orf12ng	GGYSANLFLG	190	200	210	220	230	240

-171-

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf12-1.pep	VEPQLGPGYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
	orf12ng	VEPQLGPGYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
		250	260	270	280	290	300
		310	320	330	340	350	360
10	orf12-1.pep	PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI					
	orf12ng	PETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLVI					
		310	320	330	340	350	360
		370	380	390	400	410	420
15	orf12-1.pep	IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFGILICAFINLMIGSASAQW					
	orf12ng	IFFAAQFVAFFNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFGILICAFINLMIGSASAQW					
		370	380	390	400	410	420
		430	440	450	460	470	480
20	orf12-1.pep	AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
	orf12ng	AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
		430	440	450	460	470	480
		490	500	510	520		
25	orf12-1.pep	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAPX					
	orf12ng	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTFFYPVPX					
		490	500	510	520		

In addition, ORF12ng (SEQ ID NO: 140) shows significant homology with a hypothetical protein (SEQ ID NO: 1124) from *E.coli*:

30	sp P46133 YDAH_ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION)gi 1787597 (AE000231) hypothetical protein in ogt 5' region [Escherichia coli] Length = 510 Score = 329 bits (835), Expect = 2e-89 Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)	
35	Query: 8 RSGRFLRTVEWLG NMLPHVPTXXXXXXXXXXXXASAVGAYFGLSVPDPRPVGAKGRADDGL 67 +SG+ VE +GN +PHP +A+ + FG+S +P D	
	Sbjct: 13 QSGKLYGWVERIGNKVPHPFLFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64	
40	Query: 68 IHVVSLLDADGLIKILTHTVKNFTGFAPXXXXXXXXXXXXIAEKSGSLISALMRLLLTKSP 127 + V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + +	
	Sbjct: 65 VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124	
	Query: 128 RKLTTFMVVFTGILSNTASELG YVVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187 + ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL	
	Sbjct: 125 ARYASYMVLFIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVCGGFTANL 184	
45	Query: 188 FLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247 + T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG	
	Sbjct: 185 LIVTTDVLLSGISTEAAAFAFPQMHSVSDIDNWFMASSVVVLTIVGGLITDKIIEPRLGQ 244	

Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVVALSALLAWSIVPADGILRHPETGLVA 307
 +Q + ++ + + S GL AGVV + A +A ++P +GILR P V
 Sbjet: 245 WQGNSDEKLQTLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298

Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLXXXXXXXXXX 367
 SPF+K IV I L F + + YG TR++R + ++ + M E M + ++
 Sbjet: 299 PSPFIKGIVPLIILFFVVSLAYGIATRTIRROADLPHLMIEPMKEMAGFIVMVFLAQQF 358

Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQWAVTAPIF 427
 NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF
 Sbjet: 359 VAMFNWSNMKGFIAGVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418

Query: 428 VPMLMLAGYAPEVIQAAAYRIGDSVTNIITPMSYFGLIMATVIKYKKDAGVGTLSMMLP 487
 VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP
 Sbjet: 419 VPMFMLLGFHPAFAQILFRIADSSVLPLAPVSPFVPLFLGFLQRYKPDALGTYYSLVLP 478

Query: 488 YSAFFLIAWIALFCIWVFLGLPVGPG 514
 Y FL+ W+ + W +++GLP+GPG
 Sbjet: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

20 Example 17

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 141>] (SEQ ID NO: 141):

1 ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA
 51 GGTGGAGGTG TTCGGAACA TCCAGACCGC AGTGGAAACA GGTTTTTTTC
 101 ATGGCATTTC GGTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
 151 ATgGCTTCGC GcAGTGCGTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
 201 GACGGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTgTT
 251 CTTcAGACGG CAGCAGGTcG GTTTTGTtGT ACACCTTgAT GCACGGAaTA
 301 TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
 351 TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
 401 gCGCGGTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCGGC
 451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCGGG
 501 ACT..

35 This corresponds to the amino acid sequence [<SEQ ID 142; ORF14>] (SEQ ID NO: 142; ORF14):

1 ..TAGAAGXXVF VFVTDsQVEV FGNIQTAVET GFFHGISVSS VFGAAQDSA
 51 MASRSASIPV FSATEMRtAA IFPAASRHMP VFCSSDGSRS VLLYTLMHGI
 101 SPAWISCSTF STSSICPLF GAAASTCCSS TSACAVSSSV AEKAEISLCG
 151 RXLTNPtVSV RIMLHSG..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 (SEQ ID NO: 142) shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) (SEQ ID NO: 144) from strain A of *N. meningitidis*:

```

5
    orf14.pep                                10          20          30
                                           TAGAAGXXVVFVFTDSQVEVFGNIQTAVET
                                           |:||||| |||||:|:|:|:|:|
    orf14a      GRQLGFLRVGGALFVITAQARVNNALCDCLTTGAAGFAVFVFTDGQMQVFGNVQPAVET
                  150          160          170          180          190          200

10
                  40          50          60          70          80          90
    orf14.pep      GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS
                  |||
    orf14a      GFFHGISVSSVFGAAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS
                  210          220          230          240          250          260

15
                  100          110          120          130          140          150
    orf14.pep      VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSSVAEKAEISLCG
                  |||
    orf14a      VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSSVAEKAEISLCG
                  270          280          290          300          310          320

20
                  160
    orf14.pep      RXLTNPTVSVRIMLHSG
                  |||
    orf14a      RSLTNPTVSVRIMLHSGLMYSRRRAVVSSVAKSWSFAYMPDLVSRLNRLDLP TLVX
                  330          340          350          360          370          380

```

The complete length ORF14a nucleotide sequence [<SEQ ID 143>] (SEQ ID NO: 143) is:

	1	ATGGAGGATT	TGCAGGAAAT	CGGGTTCGAT	GTCGCCGCCG	TAAAGGTAGG
	51	TCGGCAGCGC	GAACATCATC	GTCTGCATCA	TCCCCAGCCC	GGCAACGGCG
30	101	AGGCGGACGA	TGTATTGTTT	GCGTTCCTTT	TGGTTGGCGG	CTTCGATTTT
	151	TTGCGCGTCA	TAGGGTGCGG	CGGTGTAGCC	TATCTGCCTG	ATTTTCAACA
	201	GAATGTCCGA	AAGGCGGATT	TTGCCGTCTG	CCCAGACGAC	GCGGCGACGG
	251	TGCGTGCTGT	AATTGAGGTC	GATGCGGACG	ATGCCGTCTG	TACGCAAAAG
	301	CTGCTGTTTC	ATCAGCCAGA	CGCAGGCGGC	GCAGGTGATG	CCGCCGAGCA
35	351	TTAAAACCGC	CTCGCGCGTG	CCGCCGTGGG	TTTCCACAAA	GTCCGACTGG
	401	ACTTCGGGCA	GGTCGTACAG	GCGGATTTGG	TCGAGGATTT	CTTGGGGCGG
	451	CAGCTCGGTT	TTTTGCGCGT	CGGCGGTGCG	TTGTTGTGTA	TAACTGCCCC
	501	AGCCCCGCTC	AATAATGCTT	TGTGCGACTG	CCTGACAACC	GGCGCAGCAG
	551	GTTTCGCGGT	CTTCGTTTTT	GTAACGGACG	GTCAGATGCA	GGTTTTTCGGG
	601	AACGTCCAGC	CCGCAGTGA	AACAGGTTTT	TTTCATGGCA	TTTCGGTTTC
40	651	GTCTGTGTTT	GGTGCGGCGG	CACAATACTC	GGCAATGGCT	TCGCGCAGTG
	701	CGCTCTATAC	GGTATTTTCA	GCAACGGAAA	TGCGGACGGC	GGCAATTTTT
	751	CCGCGACGCT	CGGCCATAT	GCCCGTGTGT	TGTTCTTCAG	ACGCGACGAG
	801	GTCGGTTTTG	TTGTACACCT	TGATGCACGG	AATATCGCCG	GCATGGATTT
	851	CTTGCAGTAC	GTTTTCCACG	TCTTCAATCT	GCTGTCCGCT	GTTCGGAGCG
45	901	GCGGCATCGA	CGACGTGCAG	CAGCACATCG	GCTTGCGCGG	TTTCTTCCAG
	951	CGTGGCGGAA	AAGGCGGAAA	TCAGTTTGTG	CGGCAGATCG	CTGACGAATC
	1001	CGACGGTATC	GGTCAGGATA	ATGCTGCATT	CGGGACTGAT	GTACAGCCCG
	1051	GCGCGCGTCG	TGTCGAGTGT	GCGGAAAAGC	TGGTCTTTCG	CATATATGCC

1101 CGACTTGGTC AGCCGGTTGA ACAGACTGGA TTTGCCGACA TTGGTATAG

This encodes a protein having amino acid sequence [<SEQ ID 144>] (SEQ ID NO: 144):

```

5      1  MEDLQEIGFD VAAVKVGRQR EHHRLHHPQP GNGEADDVLF AFFLVGGFDF
      51  LRVIGCGGVA YLPDFQQNVG KADFAVVPDD AAARAVIEV DADDAVCTQK
     101  LLFDQPDAGG AGDAAEH*NR LARAAVGFHK VGLDFGQVVQ ADLVEDFLGR
     151  QLGFLRVGGA LRVITAQARV NNALCDCLTT GAAGFAVFVF VTDGQMVFVG
     201  NVQPAVETGF FHGISVSSVF GAAQYSAMA SRSASIPVFS ATEMRTAAIF
     251  PAASRHMPVF CSSDGSRSVL LYTLMHGISP AWISCSTFST SSICCPLFGA
10    301  AASTCSSTS ACAVSSVAE KAEISLCGRS LTNPTVSVRI MLHSGLMYSR
     351  RAVVSSVAKS WSFAYMPDLV SRLNRLDLPT LV*

```

It should be noted that this sequence includes a stop codon at position 118.

Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF14 (SEQ ID NO: 142) shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) (SEQ ID NO: 146) from *N. gonorrhoeae*:

```

      orf14.pep                                TAGAAGXXVFVFTDSQVEVFGNIQTAVET 30
      orf14ng      GRQFGFFRVGGASVITAQAGIDDALCDCLTADAAGFAVFAFVADGQMVFVGNVQPAVET 208
20    orf14.pep      GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS 90
      orf14ng      GFFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS 268
25    orf14.pep      VLLYTLMHGISPAWISCSTFSTSSICCPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG 150
      orf14ng      VLLYTLMHGISWAWISCSTFSTSSICCPLFRAAASTTCSSTSACTVSSKVAEKAEISLCG 328
      orf14.pep      RXLTNPTVSVRIMLHSG 167
      orf14ng      RSLTNPTVSVRIMLHAGLMYSRRRAVVSRAKSSWSFAYMPDLVSRLNRLDLPTLV 382

```

30 The complete length ORF14ng nucleotide sequence [<SEQ ID 145>] (SEQ ID NO: 145) is predicted to encode a protein having amino acid sequence [<SEQ ID 146>] (SEQ ID NO: 146):

```

35    1  MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
      51  LRVIGCGGVA CLPDFQQNVG EADFAVVPDD AAARAVIEV DADDAVCAQK
     101  LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
     151  QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMVFVG
     201  NVQPAVETGF FHGISVSSVF GAAQYSAMA SRSASIPVFS ATEMRTAAIF
     251  PAASRHMPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLFRA
     301  AASTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
40    351  RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*

```

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 18

- 5 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 147>] (SEQ ID NO: 147):

```

1  ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
51  GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTTATGTG ATGATTTTGA
101  TGCCGAACTC GGGCAGCTTC GGTTTCGGCT ATGCGTCGCT GGCGGCTTTG
151  TCGTTCGGCG CGCTGATGAT TGCGCTGTTA GACGTGTCGT CAAATATGGC
201  GATGCAGCCG TTAAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
251  AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
301  GTGGCGGCGA TTCTGCCGTT TGTGTTTGCG TATATCGGTT TGGCGAACAC
351  CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG
15  401  GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
451  GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCGC
501  GAATCAGGAA AAAGCCAACT GGATCGCACT CTTAAAA.CC GCGC..

```

- 20 This corresponds to the amino acid sequence [<SEQ ID 148; ORF16>] (SEQ ID NO: 148; ORF16):

```

1  ..GHYSDRTWKP RLXGRRPLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL
51  SFGALMIALL DVSSNMAMQP FKMMVGDMMN EEQKXYAYGI QSFLANTGAV
101  VAAILPFVFA YIGLANTAXK GVVPTQTVVVA FYVGAALLVI TSAFTIFKVK
25  151  EYXPETYARY HGIDVAANQE KANWIALLLX A..

```

Further work revealed the complete nucleotide sequence [<SEQ ID 149>] (SEQ ID NO: 149):

```

1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
51  AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
30  101  CCTTTACCTT GCAAAGCTCG CAAATGAGCC GCATTTTTC AAGCTAGGC
151  GCAGACCCGC ACAATTTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG
201  GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
251  CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301  GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
35  351  CTATGCGTCG CTGGCGGCTT TGTCGTTTCG CGCGCTGATG ATTGCGCTGT
401  TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAAGAT GATGGTCGGC
451  GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501  CTTAGCAAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
551  CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
601  GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
40  651  GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAACCC TACGCCCGTT
701  ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
751  CTCTTGAAAA CCGCGCCTAA GCGGTTTTGG ACGGTTACTT TGGTGCAATT
801  CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGGCA
851  TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
45  901  GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
951  GGTGATTGTG TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG

```

-176-

5
1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
1101 CATCGCTTGG GCGGGCATTG TCACTTATCC GCTGACGATT GTGACCAACG
1151 CCTTGTCGGG CAAGCATATG GGCACTTACT TGGGCTTGTT TAACGGCTCT
1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGTCGTCC
1301 TGCTGCTGGG CGCGTTTTC GTGTTCTGA TTAAAGAAAC ACACGGCGGG
1351 GTTTGA

10 This corresponds to the amino acid sequence [<SEQ ID 150; ORF16-1>] (SEQ ID NO: 150; ORF16-1):

15
20
1 MSEYTPQTAQ QGLPALAKST IWMLSFGLG VQTAFTLQSS QMSRIFQTLG
51 ADPHNLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGRR LPYLLYGTLI
101 AVIVMILMPN SGSGFGFYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
151 DMVNEEQKGY AYGISFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
201 VVVAFYVGAA LLVITSAFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYQ
301 EAGNWyGVL AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNCS
401 ICMPQIVASL LSFVLPFMLG GLQATMFLVG GVVLLLGAFS VFLIKETHGG
451 V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF16 (SEQ ID NO: 148) shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) (SEQ ID NO: 152) from strain A of *N. meningitidis*:

30
orf16.pep
orf16a
50 60 70 80 90 100
10 20 30
GHYSDRTWKPRXLGRRLLPYLLYGTLIIV
IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPRLGRRLLPYLLYGTLIIV
35
orf16.pep
orf16a
110 120 130 140 150 160
40 50 60 70 80 90
MILMPNSGSGFGFYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI
MILMPNSGSGFGFYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKGYAYGI
40
orf16.pep
orf16a
170 180 190 200 210 220
100 110 120 130 140 150
QSFLANTGAVVAAILPFVFAVIGLANTAXKGVVPQTVVVAFYVGAAALLVITSAFTIFKVK
QSFLANTGAVVAAILPFVFAVIGLANTAEKGVVPQTVVVAFYVGAAALLVITSAFTIFKVK
45
orf16.pep
orf16a
230 240 250 260 270 280
160 170 180
EYXPETYARYHGIDVAANQEKANWIALLKXA
EYNPETYARYHGIDVAANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI

orf16a AENVVHTTDASSVG YQEAGN WYGV LAAVQS VAAVICS FVLAKVP NKYHKAGY FGCLALGA
290 300 310 320 330 340

The complete length ORF16a nucleotide sequence [<SEQ ID 151>] (SEQ ID NO: 151) is:

```

5      1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
      51  AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
     101  CCTTTACCCCT GCAAAGCTCG CAGATGAGCC GCATCTTCCA GACGCTCGGT
     151  GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG
     201  GATGCTGGTG CAGCGATTG TCGGCCATTA CTCCGACCGC ACTTGGGAAGC
    10    251  CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
      301  GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
      351  CTATGCGTCG CTGGCGGCTT TGTGCTTCGG CGCGCTGATG ATTGCGCTGT
      401  TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGTCGGC
      451  GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
    15    501  CTTAGCGAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
      551  CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
      601  GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
      651  GTTCACGATT TTCAAAGTGA AGGAATACAA TCCGAAAACC TACGCCCGTT
      701  ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
    20    751  CTCTTGAAAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGCAATT
      801  CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
      851  TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCGT AGGTTATCAG
      901  GAGGCGGGTA ACTGGTACGG CGTTTGGCG GCGGTGCAGT CGGTGCGGCG
      951  GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
    25   1001  CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
     1051  TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTAATCGG
     1101  CATCGCTTGG GCGGGCATT AACTTATCC GCTGACGATT GTGACCAACG
     1151  CCTTGTCGGG CAAGCATATG GGCATTACT TGGGCCTGTT TAACGGCTCT
     1201  ATCTGTATGC CCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
    30   1251  TATGCTGGGC GGCTTGCAAG CCACTATGTT CTTGGTAGGG GCGCTCGTCC
      1301  TGCTGCTGGG CGCGTTTCC GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG
      1351  GTTTGA
  
```

This encodes a protein having amino acid sequence [<SEQ ID 152>] (SEQ ID NO: 152):

```

35      1  MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
      51  ADPHSLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI
     101  AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
     151  DMVNEEQGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVPQT
     201  VVVAFYVGAA LLVITSFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE
    40    251  LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVG YQ
      301  EAGN WYGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
      351  FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNGS
     401  ICMPQIVASL LSFVLPMLG GLQATMFLVG GVVLLLGAFS VFLIKETHGG
    45   451  V*
  
```

ORF16a (SEQ ID NO: 152) and ORF16-1 (SEQ ID NO: 150) show 99.6% identity in 451 aa overlap:

```

50      orf16a .pep      10      20      30      40      50      60
                        MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF
                        |||||
      orf16-1          MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFF
                        10      20      30      40      50      60
  
```

-178-

5	orf16a.pep	70 80 90 100 110 120	ILPPLAGMLVQPIVGHYSDRTWKPRLLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS
	orf16-1	70 80 90 100 110 120	ILPPLAGMLVQPIVGHYSDRTWKPRLLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS
10	orf16a.pep	130 140 150 160 170 180	LAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQGYAYGIQSFLANTGAVVAAILP
	orf16-1	130 140 150 160 170 180	LAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQGYAYGIQSFLANTGAVVAAILP
15	orf16a.pep	190 200 210 220 230 240	FVFAYIGLANTAEGVVPQTVVVAFYVGAALLVITSFTIFKVKEYNPETYARYHGIDVA
	orf16-1	190 200 210 220 230 240	FVFAYIGLANTAEGVVPQTVVVAFYVGAALLVITSFTIFKVKEYDPETYARYHGIDVA
20	orf16a.pep	250 260 270 280 290 300	ANQEKANWIELLKTAPKAFWTVTLVQFFCWFQYMWTSAGAIENVWHTTDASSVGYQ
	orf16-1	250 260 270 280 290 300	ANQEKANWIELLKTAPKAFWTVTLVQFFCWFQYMWTSAGAIENVWHTTDASSVGYQ
25	orf16a.pep	310 320 330 340 350 360	EAGNWWYGVLAQVSVAAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV
	orf16-1	310 320 330 340 350 360	EAGNWWYGVLAQVSVAAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV
30	orf16a.pep	370 380 390 400 410 420	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNIGSICMPQIVASLLSFVLPMLG
	orf16-1	370 380 390 400 410 420	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNIGSICMPQIVASLLSFVLPMLG
35	orf16a.pep	430 440 450	GLQATMFLVGGVLLLGAFSVFLIKETHGGVX
	orf16-1	430 440 450	GLQATMFLVGGVLLLGAFSVFLIKETHGGVX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF16 (SEQ ID NO: 148) shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) (SEQ ID NO: 154) from *N. gonorrhoeae*:

40	orf16.pep	GHYSDRTWKPRLLGGRRLPYLLYGTLIAVIV	30
	orf16ng	HFSNARRRRPAQFGLVFHPAAAGDAGSADSGYYSDRTWKPRLLGGRRLPYLLYGTLIAVIV	131
45	orf16.pep	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKXYAYGI	90
	orf16ng	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKSYAYGI	191
45	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSFTIFKVK	150

	:	
orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAALLIITSFTISKVK	251
orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
	:	
5 orf16ng	EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWTVTTPVQFFCWFAFRYMWTYSAGAI	311

The complete length ORF16ng nucleotide sequence [<SEQ ID 153>] (SEQ ID NO: 153) is:

	1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
10	51	TACTTTTCAA	ATCAAAAAAA	AGGATTTACT	TTATGTCGGA	ATATACGCCT
	101	CAAACAGCAA	AACAAGGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTTGGAT
	151	GTTGAGCTTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAGCT
	201	CGCAGATGAG	CCGCATTTTT	CAAACGCTAG	GCGCAGACCC	GCACAATTTG
	251	GGCTGGTTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG	TTACGCCGAT
15	301	AGTGGCTACT	ACTCAGACCG	CAC TTGGAAG	CCGCGCTTGG	GCGGCCGCCG
	351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTCATC	GTGATGATTT
	401	TGATGCCGAA	CTCGGGCAGC	TTCGGTTTCG	GCTATGCGTC	GCTGGCGGCC
	451	TTGTCGTTCG	GCGCGCTGAT	GATTGCGCTG	TTGGACGTGT	CGTGAATAT
	501	GGCGATGCAG	CCGTTTAAAG	TGATGGTCGG	CGATATGGTC	AACGAGGAGC
20	551	AGAAAAGCTA	CGCCTACGGG	ATTCAAAGTT	TCTTAGCGAA	TACGGACGCG
	601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTTT	CGTATATACG	GTTTGGCGAA
	651	CATTGCCGAG	AAAGGCGTTG	TGCCACAAC	CGTGATTCGTA	GCATTCTATG
	701	TGGGTGCGGC	GTTACTGATT	ATTACCAAGT	CGTTCACAAT	CTCCAAAGTC
	751	AAAGAATACG	ACCCGGAAC	CTACGCCCGT	TACCACGGCA	TCGATGTCGC
25	801	CGCGAATCAG	GAAAAAGCCA	ACTGGTTTCA	ACTCTTAAAA	ACCGCGCCTA
	851	AAGTGTTTTG	GACGGTACT	CCGGTACAGT	TTTTCTGCTG	GTTCGCCTTC
	901	CGGTATATGT	GGACTTACTC	GGCAGGCGCG	ATTGCAGAAA	ACGTCTGGCA
	951	CACTACCGAT	GCGTCTTCCG	TAGGCCATCA	GGAGCGGGC	AACCGGTACG
	1001	GCGTTTTGGC	GGCGGTGTAG			

30 This encodes a protein having amino acid sequence [<SEQ ID 154>] (SEQ ID NO: 154):

35

1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLLLYVG	IYASNSKTRF	ARAGKKHDL
51	VELRLSRSD	GLYPAKLAD	PHFSNARRRP	AQFGLVFHPA	AAGDGAGSAD
101	SGYSDRTWK	PRLGGRRLPY	LLYGTLI AVI	VMIL M PNS SGS	FGFGYAS LAA
151	LS F GAL MIAL	LD VSSNMAMQ	PFKMMVGDMV	NEEQKSAYAG	IQSFLANTDA
201	VVAAILPFVF	AYIGLANTAE	KGVPQTVVW	AFYVGAALLI	ITSFTISKV
251	KEYDPETYAR	YHIDVAA NQ	EKANWFELLK	TAPKVFVTVT	PVQFFCWPAF
301	RYMW TS SAGA	IAENV W HTTD	ASSVGHQ EAG	NRYGV LAA V*	

ORF16ng (SEQ ID NO: 154) and ORF16-1 (SEQ ID NO: 150) show 89.3% identity in 261 aa

40 overlap:

		30	40	50	60	70	80
	orf16-1.pep	MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSVRT					
					: : :		:
45	orf16ng	DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGGDAGSADSGYYSVRT					
		50	60	70	80	90	100
		90	100	110	120	130	140
	orf16-1.pep	WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSGFGFYASLAALSFGALMIALLDVSSNMA					
50	orf16ng	WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSGFGFYASLAALSFGALMIALLDVSSNMA					
		110	120	130	140	150	160

-180-

		150	160	170	180	190	200
	orf16-1.pep	MQPFKMMVGDMVNEEQGYAYGIQSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTV					
	orf16ng	MQPFKMMVGDMVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTV					
5		170	180	190	200	210	220
		210	220	230	240	250	260
	orf16-1.pep	VVAFYVGAALLVITSAFTIFKVKEYDPETYARYHGIDVAANQEKANWIELLKTA PKAFWT					
	orf16ng	VVAFYVGAALLIITSAFTISKVKEYDPETYARYHGIDVAANQEKANWFELLKTA PKVFWT					
10		230	240	250	260	270	280
		270	280	290	300	310	320
	orf16-1.pep	VTLVQFFCWFQYMWTYSAGAIAENVWHTTDASSVGQEAGNWWYGVLA AVQSVAAVICS					
	orf16ng	VTPVQFFCWFQYFRYMWWTYSAGAIAENVWHTTDASSVGHQEAGNRYGVLA AVX					
15		290	300	310	320	330	340

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 19

20 The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 155](#)] ([SEQ ID NO: 155](#)):

25

```

1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGCATA CCTTGATGCT
51  GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGCGAAA
101 CAATCACCCG NAAACACGTT GNCAAAGACC AAATCCGNGN CTCGGTGTG
151 GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201 CGGAAAATAC TGGTTCGTCG TCAATCCCGA AGATTCGGCG AA.NTGACGG
251 GNATTTTGAN GGCAGGGCTG GACAAACCTT TCCAAATAGT TNAGGATACC
301 CCGAGCTATG C.TGCCACCA AGCCCTGCCG GTCAAACTCG GATCGNCTGG
351 CAGCCAGAAT...
```

30 This corresponds to the amino acid sequence [[SEQ ID 156; ORF28](#)] ([SEQ ID NO: 156; ORF28](#)):

35

```

1  MLFRKTTAAV LAHTLMLNGC TLMLWGMNNP VSETITRKHV XKDQIRXFGV
51  VAEDNAQLEK GSLVMMGGKY WFVVPEDSA XXTGILXAGL DKPFQIVXDT
101 PSYXCHQALP VKLGSXGSQN...
```

Further work revealed the complete nucleotide sequence [[SEQ ID 157](#)] ([SEQ ID NO: 157](#)):

40

```

1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATGCT
51  GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGCGAAA
101 CAATCACCCG CAAACACGTT GACAAAGACC AAATCCGCGC CTCGGTGTG
151 GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201 CGGAAAATAC TGGTTCGTCG TCAATCCCGA AGATTCGGCG AAGCTGACGG
```

5	251	GCATTTTGAA	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TGAGGATACC
	301	CCGAGCTATG	CTCGCCACCA	AGCCCTGCCG	GTCAAACCTG	AATCGCCTGG
	351	CAGCCAGAAT	TTCAGTACCG	AAGGCCTTTG	CCTGCGCTAC	GATACCGACA
	401	AGCCTGCCGA	CATCGCCAAG	CTGAAACAGC	TCGGGTTTGA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAAATA
	501	CTACGCCACA	CCGCAAAAAC	TGAACGCCGA	TTACCATTTT	GAGCAAAGTG
	551	TGCCTGCCGA	TATTTATTAC	ACGGTTACTG	AAGAACATAC	CGACAAATCC
	601	AAGCTGTTTG	CAAATATCTT	ATATACGCCC	CCCTTTTTGA	TACTGGATGC
	651	GGCGGGCGCG	GTACTGGCCT	TGCCTCGGGC	GGCTCTGGGT	GCGGTCGTGG
	701	ATGCCGCCCG	CAAATGA			
10						

This corresponds to the amino acid sequence [[<SEQ ID 158; ORF28-1>](#)] ([SEQ ID NO: 158; ORF28-1](#));

15

1	MLFRKTTAAV	LAATLMLNGC	TLMLWGMNPN	VSETITRKHV	DKDQIRAFGV
51	VAEDNAQLEK	GSLVMMGGKY	WVVNPEDSA	KLTGILKAGL	DKPFQIVEDT
101	PSYARHQALP	VKLESPGSQN	FSTEGLCRLY	DTDKPADIAK	LKQLGFQAVK
151	LDNRNTIYTRC	VSAGKYYAT	PQKLNADYHF	EQSPADIYY	TVTEEHTDKS
201	KLFANILYTP	PFLILDAAGA	VLALPAAALG	AVVDAARK*	

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 (SEQ ID NO: 156) shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) (SEQ ID NO: 160) from strain A of *N. meningitidis*:

			10	20	30	40	50	60
25	orf28.pep	<u>MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK</u>						
	orf28a	<u>MLFRKTTAAVLAATLMLNGCTVMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK</u>						
			10	20	30	40	50	60
30	orf28.pep	GSLVMMGGKYWFVVPNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXGSQN						
	orf28a	GSLVMMGGKYWFVVPNPEDSAKLTGILKAGLDKQFQMVEPNPRFA-YQALPVKLESASPQN						
			70	80	90	100	110	120
			70	80	90	100	110	
35	orf28a	FSTEGLCRLRYDTRDPADIAKLKQLEFEAVELDNRTIYTRCVSAGKGYATPQKLNADYHF						
			120	130	140	150	160	170

The complete length ORF28a nucleotide sequence [<SEQ ID 159>] (SEQ ID NO: 159) is:

	1	ATGTTGTTCC	GTA AACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGTT
	51	GAACGGCTGT	ACG GTAATGA	TGTGGGGTAT	GAACAGCCCG	TTCAGCGAAA
40	101	CGACCGCCCG	CAAACACGTT	GACAAGGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGA AAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGGAAATAC	TGGTTCGTCG	TCAATCCTGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCCGGGTTG	GACAAGCAGT	TTCAAATGGT	TGAGCCCAAC
	301	CCGCGCTTTG	CCTACCAAGC	CCTGCCGGTC	AAACTCGAAT	CGCCCGCCAG
	351	CCAGAATTTC	AGTACCGAAG	GCCTTTGCCT	GCGCTACGAT	ACCGACAGAC
45	401	CTGCCGACAT	CGCCAAGCTG	AAACAGCTTG	AGTTTGAAAG	GGTGC AA CTC
	451	GACAATCGGA	CCATTTTACAC	GCGCTGCGTC	TCCGCCAAAG	GCAAATACTA

501 CGCCACACCG CAAAACTGA ACGCCGATTA TCATTTTGAG CAAAGTGTGC
 551 CTGCCGATAT TTATTACACG GTTACGAAAA AACATACCGA CAAATCCAAG
 601 TTGTTTGAAA ATATTGCATA TACGCCACC ACGTTGATAC TGGATGCGGT
 651 GGGCGCGGTG CTGGCCTTGC CTGTCGCGGC GTTGATTGCA GCCACGAATT
 701 CCTCAGACAA ATGA

This encodes a protein having amino acid sequence [<SEQ ID 160>] (SEQ ID NO: 160):

1 MLFRKTTAAV LAATLMLNGC TVMMWGMNSP FSETTARKHV DKDQIRAFGV
 51 VAEDNAQLEK GSLVMMGGKY WVVNPEDSA KLTGILKAGL DKQFQMVEPN
 101 PRFAYQALPV KLESPASQNF STEGLCLRYD TDRPADIACL KQLEFEAVEL
 151 DNRTIYTRCV SAKGKYYATP QKLNADYHFE QSVPADIIYTT VTKKHTDKSK
 201 LFENIAYTPT TLILDAVGAV LALPVAALIA ATNSSDK*

ORF28a (SEQ ID NO: 160) and ORF28-1 (SEQ ID NO: 158) show 86.1% identity in 238 aa overlap:

	10	20	30	40	50	60
orf28a.pep	MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK					
orf28-1	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
	10	20	30	40	50	60
orf28a.pep	GSLVMMGGKYWVVNPEDSAKLTGILKAGLDKQFQMVEPNPRFA-YQALPVKLESPASQN					
orf28-1	GSLVMMGGKYWVVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN					
	70	80	90	100	110	119
orf28a.pep	GSLVMMGGKYWVVNPEDSAKLTGILKAGLDKQFQMVEPNPRFA-YQALPVKLESPASQN					
orf28-1	GSLVMMGGKYWVVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN					
	70	80	90	100	110	120
orf28a.pep	120	130	140	150	160	170
orf28a.pep	FSTEGCLCLRYDTDRPADIACLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF					
orf28-1	FSTEGCLCLRYDTDKPADIACLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
	130	140	150	160	170	180
orf28a.pep	180	190	200	210	220	230
orf28a.pep	EQSVPADIIYTVTKKHTDKSKLFENIAYTPTTLILDAVGAVLALPVAALIAATNSSDKX					
orf28-1	EQSVPADIIYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
	190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF28 (SEQ ID NO: 156) shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) (SEQ ID NO: 162) from *N. gonorrhoeae*:

orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK	60
orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKDQIRAFGVVAEDNAQLEK	60
orf28.pep	GSLVMMGGKYWVVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXSQN	120
orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN	120

The complete length ORF28ng nucleotide sequence [<SEQ ID 161>] (SEQ ID NO: 161) is

```

1  ATGTTGTTC  GTAAACGAC  CGCCGCCGTT  TTGGCGGCAA  CCTTGATACT
5  51  GAACGGCTGT  ACGATGATGT  TGCGGGGGAT  GAACAACCCG  GTCAGCCAAA
101 CAATCACCCG  CAAACACGTT  GACAAAGACC  AAATCCGCGC  CTTCGGTGTG
151 GTTGCCGAAG  ACAATGCCCA  ATTGGAAAAG  GGCAGCCTGG  TGATGATGGG
201 CGGGAAATAC  TGGTTCGCCG  TCAATCCCGA  AGATTCGGCG  AAGCTGACGG
251 GCCTTTTGAA  GGCCGGGTTG  GACAAGCCCT  TCCAAATAGT  TGAGGATACC
301 CCGAGCTATG  CCCGCCACCA  AGCCCTGCCG  GTCAAATTCG  AAGCGCCCGG
10 351 CAGCCAGAAT  TTCAGTACCG  GAGGTCTTTG  CCTGCGCTAT  GATACCGGCA
401 GACCTGACGA  CATCGCCAAG  CTGAAACAGC  TTGAGTTTAA  AGCGGTCAAA
451 CTCGACAATC  GGACCATTTA  CACGCGCTGC  GTATCCGCCA  AAGGCAAATA
501 CTACGCCACG  CCGCAAAAAC  TGAACGCCGA  TTATCATTTT  GAGCAAAGTG
551 TGCCCGCCGA  TATTTATTAT  ACGGTTACTG  AAAAAATAC  CGACAAATCC
15 601 AAGCTGTTTG  GAAATATCTT  ATATACGCCC  CCCTTGTTGA  TATTGGATGC
651 GCGGCCGCG  GTGCTGGTCT  TGCCTATGGC  TCTGATTGCA  GCCGCGAATT
701 CCTCAGACAA  ATGA

```

This encodes a protein having amino acid sequence [<SEQ ID 162>] (SEQ ID NO: 162):

```

20 1  MLFRKTTAAV  LAATLILNGC  TMMLRGMNPN  VSQTITRKHV  DKDQIRAFGV
51  VAEDNAQLEK  GSLVMMGGKY  WFAVNPEDSA  KLTGLLKAGL  DKPFQIVEDT
101 PSYARHQALP  VKFEAPGSQN  FSTGGLCLRY  DTGRPDDIAK  LKQLEFKAVK
151 LDNRTIYTRC  VSAKGKYYAT  PQKLNADYHF  EQSVPADIYY  TVTEKHTDKS
25 201 KLFGNILYTP  PLLILDAAAA  VLVLPMALIA  AANSSDK*

```

ORF28ng (SEQ ID NO: 162) and ORF28-1 (SEQ ID NO: 158) share 90.0% identity in 231 aa overlap:

```

30 orf28-1.pep  10      20      30      40      50      60
    MLFRKTTAAVLAATLMLNGCTLMWGMNPNVSETITRKHVVDKDQIRAFGVVAEDNAQLEK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf28ng  MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVVDKDQIRAFGVVAEDNAQLEK
    10      20      30      40      50      60

35 orf28-1.pep  70      80      90      100     110     120
    GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf28ng  GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN
    70      80      90      100     110     120

40 orf28-1.pep  130     140     150     160     170     180
    FSTEGCLCLRYDTPADIAKLKQLGFQAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf28ng  FSTGGLCLRYDTPGRPDDIAKLKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF
    130     140     150     160     170     180

45 orf28-1.pep  190     200     210     220     230     239
    EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf28ng  EQSVPADIYYTVTEKHTDKSKLFGNILYTPPLLILDAAAVALVLPMALIAAANSSDKX
    190     200     210     220     230

```

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (SEQ ID NO: 158) (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 (SEQ ID NO: 158) is a surface-exposed protein, and that it may be a useful immunogen.

Example 20

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 163] (SEQ ID NO: 163):

```

15      1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
      51  TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
     101  CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTTCAG CGGCGGTGTA
     151  GACGGCGGTT TTACTGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
     201  TCCGAGGAGT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCGG
     251  GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
20     301  ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
     351  AGAAAATGCC GGTGCCGCCT CTGGT..

```

This corresponds to the amino acid sequence [SEQ ID 164; ORF29] (SEQ ID NO: 164; ORF29):

```

25      1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHDS KSTSDFSGGV
      51  DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYVVKGTSTK
     101  TKTSIVPQAP FSDRWLEENA GAASG..

```

Further work revealed the complete nucleotide sequence [SEQ ID 165] (SEQ ID NO: 165):

```

30      1  ATGAATTGTC CTATTCAAAA ATTCATGATG CTGTTTGACG CAGCAATATC
      51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
     101  GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
     151  TTTGGTAATG CTCGCGGCAG TGTAAAAAG CGGGTTTACG CCGTCCAGAC
     201  ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
35     251  CAGGGTTTGA AGGTGTTATC GGTATGAAA CCCATTTTTC AGGGCACGGA
     301  CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
     351  TTTCAGCGGC GGTGTAGACG GCGGTTTAC TGTTTACCAA CTTCATCGAA
     401  CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC

```


-185-

5 451 GATTATCCGC CCCCCGGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
 501 AGGAACCTCA ACAAAAACAA AGACTAATAT TGTCCTCAA GCCCATT
 551 CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCTCTGG TTTTTCAGC
 601 CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
 651 TTGGTGGGCT AACCGTATGG ATGATGTTTCG CGGCATCGTC CAAGGTGCGG
 701 TTAATCCTTT TTAAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
 751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
 801 AGGTATTAAT GATTTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG
 851 CGAGCCTATT ACAGGACAGT GCTTTTGC GG TAAAAGACGG TATCAACTCT
 10 901 GCCAAACAAT GGGCTGATGC CCATCCAAAT ATAAACAGCTA CTGCCCAAAC
 951 TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTTGGAGA GGTAAAAAAG
 1001 TAGAACTTAA CCCGACTAAA TGGGATTGGG TAAAAAATAC CGGTATATAA
 1051 AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
 1101 GAATAAACCT ATTAATCTT TACCAAAACAG TGCCGCTGAA AAAAGAAAAC
 15 1151 AAAATTTTGA GAAGTTTAA AGTAACTGGA GTTCAGCAAG TTTTGATTCA
 1201 GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA
 1251 AGTTAAACT CGATACACTA GTTTAGATGG AAAAATTACA ATTATAAAAG
 1301 ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCACG AAAACAGTAT
 1351 CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAATTTAC AAGGTAAGCA
 20 1401 AGCAAAAGAT TATTTACAAC AACAACTCA TATCAGGAAC TTAGACAAAT
 1451 GA

This corresponds to the amino acid sequence [SEQ ID 166; ORF29-1] (SEQ ID NO: 166; ORF29-1):

25 1 MNLPQKFM LFAAAISLLQ IPISHANGLD ARLRDMQAK HYEPPGKYHL
 51 FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG
 101 HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
 151 DYPPPGGARD IYSYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS
 201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT
 30 251 DSAVSPVTD T AAQQT LQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS
 301 AKQWADAHPN ITATAQTALS AAEEAGTVWR GKKVELNPTK WDWVKNTRYK
 351 KPAARHMQTL DGEMAGGNKP IKSLPNSAAE KRKQNFEPFN SNWSSASFDS
 401 VHKTLPNAP GILSPDKVKT RYTSLDGKIT I IKDNENNYF RIHDNSRKQY
 451 LDSNGNAVKT GNLQKQAKD YLQQQTHIRN LDK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF29 (SEQ ID NO: 164) shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) (SEQ ID NO: 168) from strain A of *N. meningitidis*:

40 orf29.pep VSPVLPIT HERTGFEGVIGYETHFSGHGHE
 orf29a EPPGKYHLFGNARGSVKNRVYAVQTFDATAVGPILPITHERTGFEGIIGYETHFSGHGHE
 50 60 70 80 90 100

45 orf29.pep VHSPFDHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDGYDGPQAAAXYPPPGGARDIY
 orf29a VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIY
 110 120 130 140 150 160

-186-

```

      100      110      120
orf29.pep  SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG
           |||||:|||:|||:|||:|||:|||:|||:|||
orf29a     XXVVKGTSTKTKSNIVPRAPFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANR
5          170      180      190      200      210      220

orf29a     MDDIRGIVQGA VNPFLMGFQGVGIGAITDSAVSPVTDTAQQTLQGXNHLGXLSPEAQLA
           230      240      250      260      270      280

```

The complete length ORF29a nucleotide sequence [<SEQ ID 167>] (SEQ ID NO: 167) is:

```

10      1  ATGAATTNGC  CTATTCAAAA  ATTCATGATG  CTGTTTGCAG  CAGCAATATC
      51  GTNGCTGCAA  ATCCCNATTA  GTCATGCGAA  CGGTTTGGAT  GCCCGTTTGC
     101  GCGATGATAT  GCAGGCAAAA  CACTACGAAC  CGGGTGGTAA  ATACCATCTG
     151  TTTGGTAATG  CTCGCGGCAG  TGTAAAAAAT  CGGGTTTACG  CCGTCCAAAC
     201  ATTTGATGCA  ACTGCGGTCTG  GCCCCATACT  GCCTATTACA  CACGAACGGA
     15      251  CAGGATTTGA  AGGCATTATC  GGTATGAAA  CCCATTTTTC  AGGACATGGA
     301  CATGAAGTAC  ACAGTCCGTT  CGATAATCAT  GATTCAAAAA  GCACTTCTGA
     351  TTTCAGCGGC  GGCCTAGACG  GTGGTTTAC  CGTTTACCAA  CTTCATCGGA
     401  CAGGGTCGGA  AATCCATCCG  GAGGATGGAT  ATGACGGGCC  GCAAGGCAGC
     451  GATTATCCGC  CCCC CGGAGG  AGCAAGGGAT  ATATACANNT  ANTATGTCAA
     20      501  AGGAACTTCA  ACAAAAAACA  AGAGTAATAT  TGTTCCCCGA  GCCCCATTTT
     551  CAGACCGCTG  GCTAAAAGAA  AATGCCGGTG  CCGCCTCTGG  TTTTTCAGC
     601  CGTGCTGATG  AAGCAGGAAA  ACTGATATGG  GAAAGCGACC  CCAATAAAAA
     651  TTGGTGGGCT  AACCGTATGG  ATGATATTCG  CGGCATCGTC  CAAGGTGCGG
     701  TTAATCCTTT  TTTAATGGGT  TTTCAAGGAG  TAGGGATTGG  GGCAATTACA
     25      751  GACAGTGCAG  TAAGCCCGGT  CACAGATACA  GCCGCGCAGC  AGACTCTACA
     801  AGGTATNAAT  CATTTAGGAA  ANTTAAGTCC  CGAAGCACAA  CTTGCGGCTG
     851  CAACCGCATT  ACAAGACAGT  GCTTTTGC GG  TAAAAGACGG  TATCAATTCC
     901  GCCAGACAAT  GGGCTGATGC  CCATCCGAAT  ATAAGTCAA  CAGCCCAAAC
     951  TGCCCTTGCC  GTAGCAGANG  CCGCAACTAC  GGTTTGGGGC  GGTAATAAAG
     30      1001  TAGAACTTAA  CCCGACCAAA  TGGGATTGGG  TTAATAATAC  NGGCTATAAN
     1051  ACACCTGCTG  TTCGCACCAT  GCATACTTTG  GATGGGGAAA  TGGCCGGTGG
     1101  GAATAGACCG  CCTAAATCTA  TAACGTCCA  CAGCAAAGCA  GATGCTTCCA
     1151  CACAACCGTC  TTTACAAGCG  CAACTAATTG  GAGAACAAAT  TANNNNNGGG
     1201  CATGCTTATA  ACAAGCATGT  CATAAGACAA  CAAGAATTTA  CGGATTTAAA
     35      1251  TATCAATTCA  CCAGCAGATT  TTGCTCGGCA  TATTGAAAAT  ATGTGTAGCC
     1301  ATCCANCAAA  TATGAAAGAG  TTACCTCGCG  GTAGAACTGC  GTATTGGGAT
     1351  NATAAACAG  GGACNATAGT  TATCCGAGAT  AAAAATTCTG  ACGATGGAGG
     1401  TACAGCATTT  AGACCAACAT  CAGGTAAAAA  ATATTATGAT  GATTTATAG

```

40 This encodes a protein having amino acid sequence [<SEQ ID 168>] (SEQ ID NO: 168):

```

      1  MNXPIQKFMM  LFAAAISXLQ  IPISHANGLD  ARLRDDMQAK  HYEPGGKYHL
     51  FGNARGSVKN  RVYAVQTFDA  TAVGPILPIT  HERTGFEGII  GYETHFSGHG
     101  HEVHSPFDNH  DSKSTSDFSG  GVDGGFTVYQ  LHRTGSEIHP  EDGYDGPQGS
     151  DYPPPGGARD  IYXXVVKGTS  TKTKSNIVPR  APFSDRWLKE  NAGAASGFFS
     45      201  RADEAGKLIW  ESDPNKNWWA  NRMDDIRGIV  QGA VNPFLMG  FQGVGIGAIT
     251  DSAVSPVTD  AAQQT LQGXN  HLGXLSPEAQ  LAAATALQDS  AFAVKDGINS
     301  ARQWADAHPN  ITATAQTALA  VAXAATTVWG  GKKVELNPTK  WDWVKNTGYX
     351  TPAVRTMHTL  DGEMAGGNRP  PKSITSNSKA  DASTQPSLQA  QLIGEQIXXG
     401  HAYNKHVIRQ  QEFTDLNINS  PADFARHIEN  IVSHPXNMKE  LPRGRTAYWD
     50      451  XKTGTIVIRD  KNSDDGGTAF  RPTSGKKYYD  DL*

```

ORF29a (SEQ ID NO: 168) and ORF29-1 (SEQ ID NO: 166) show 90.1% identity in 385 aa overlap:

-187-

		10	20	30	40	50	60
	orf29a.pep	MNXPIQKFMMLFAAAISXLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN					
5	orf29-1	MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK					
		10	20	30	40	50	60
	orf29a.pep	70	80	90	100	110	120
	orf29-1	RVYAVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG					
10		70	80	90	100	110	120
	orf29a.pep	130	140	150	160	170	180
	orf29-1	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYXXYVKGSTSTKTSNIVPR					
15		130	140	150	160	170	180
	orf29a.pep	190	200	210	220	230	240
	orf29-1	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDIRGIVQGAVNPFLMG					
20		190	200	210	220	230	240
	orf29a.pep	250	260	270	280	290	300
	orf29-1	FQGVGIGAITDSAVSPVTDTAQQTLQG XNHLGXLSPAQLAAATALQDSAFVKG DINS					
25		250	260	270	280	290	300
	orf29a.pep	310	320	330	340	350	360
	orf29-1	ARQWADAHPNITATAQTALAVAXAATT VWGKKVELNPTKWDVWKNTGYXTPAVRTMHTL					
30		310	320	330	340	350	360
	orf29a.pep	370	380	390	400	410	420
	orf29-1	DGEMAGGNRPPKSITSNSKADASTQPSLQAQLIGE QIXXGHAYNKHVIRQQEFTDLNINS					
35		370	380	390	400	410	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 (SEQ ID NO: 164) shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) (SEQ ID NO: 170) from *N. gonorrhoeae*:

40	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSGHGHE	30
	orf29ng	EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE	102
	orf29.pep	VHSPFDHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDGYDGPQAAXYPPPGGARDIY	90
	orf29ng	VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGYPPPGGARDIY	162
45	orf29.pep	SYVVKGTSTKTKTSIVQAPFSDRWLEENAGAASG	125

orf29ng ||:::||||| : |||||:|||||
 SYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANR 222

The complete length ORF29ng nucleotide sequence [[<SEQ ID 169>](#)] ([SEQ ID NO: 169](#)) is
 5 predicted to encode a protein having amino acid sequence [[<SEQ ID 170>](#)] ([SEQ ID NO: 170](#)):

1 MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPPGKGYHL
 51 FGNARGSVKN RVCVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGG
 151 GYPPPGGARD IYSYHIKGTSTKTKINTVPQ APFSDRWLKE NAGAASGFLS
 201 RADEAGKLIW ENDPDKNWR NRMDDIRGIV QGAVNPFLTG FQGLGVGAIT
 251 DSAVSPVTYA AARKTLQGIH NLGNLSPEAQ LAAATALQDS AFAVKDSINS
 301 ARQWADAHPN ITATAQTALA VTEAATTWVG GKKVELNPAK WDWVKNTGYK
 351 KPAARHMQTV DGEMAGGNKP LESKNTVTNN NFFENTGYTE KVLKQASNGD
 401 YHGFQSVDA FSENGTVIQI VGGDNIVRHK LYIPGSYKKG DGNFEYIREA
 451 DGKINHRLFV PNQQLPEK*

In a second experiment, the following DNA sequence [[<SEQ ID 171>](#)] ([SEQ ID NO: 171](#)) was
 identified:

1 atgAATTTGC CTATTCAAAA ATTCATGATG ctgttggcAg cggcaatatac
 51 gatgctGCat ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
 101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGCAA ATACCATCTG
 151 TTTGGTAATG CTCGCGGCAG TGTAAAAAAT CGGGTTTGC GCGTCCAAAC
 201 ATTTGATGCA ACTGCGGTCG GCCCCATACT GCCTATTACA CACGAACGGA
 251 CAGGATTGGA AGGTGTTATC GGCTATGAAA CCCATTTTTC AGGACACGGA
 301 CACGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA
 351 TTTCAGCGGC GCGGTAGACG CGGTTTAC CGTTTACCAA CTTCATCGGA
 401 CAGGGTCGGA AATACATCCC GCAGACGGAT ATGACGGGCC TCAAGGCGGC
 451 GGTATCCGG AACCACAAGG GGCAAGGGAT ATATACAGCT ACCATATCAA
 501 AGGAACCTCA ACCAAAAACA AGATAAACAC TGTTCGCAA GCCCCTTTT
 551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCTCCGG TTTTCTCAGC
 601 CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAACGACC CCGATAAAAA
 651 TTGGCGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG
 701 TTAATCCTTT TTTAACGGGT TTTCAAGGGG TAGGGATTGG GGCAATTACA
 751 GACAGTGCGG TAAGCCCGGT CACAGATACA GCCGCTCAGC AGACTCTACA
 801 AGGTATTAAT GATTTAGGAA ATTTAAGTCC GGAAGCACAA CTTGCCGCCG
 851 CGAGCCTATT ACAGGACAGT GCCTTTGCGG TAAAAGACGG CATCAATTCC
 901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACAGCAA CAGCCCAAAC
 951 TGCCCTTGCC GTAGCAGAGG CCGCAGGTAC GGTTTGGCGC GGTAAAAAAG
 1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TAAAAATAC CGGCTATAAA
 1051 AAACCTGCTG CCCGCCATAT GCAGACTGTA GATGGGGAGA TGGCAGGGGG
 1101 GAATAGACCG CCTAAATCTA TAACGTCGGA AGGAAAAGCT AATGCTGCAA
 1151 CCTATCCTAA GTTGGTTAAT CAGCTAAATG AGCAAACTT AAATAACATT
 1201 GCGGCTCAAG ATCCAAGATT GAGTCTAGCT ATTCATGAGG GTAAAAAATA
 1251 TTTTCCAATA GGAAGTCAA CTTATGAAGA GGCAGATAGA CTAGGTAAAA
 1301 TTTGGGTTGG TGAGGGTGCA AGACAACTA GTGGAGGCGG ATGGTTAAGT
 1351 AGAGATGGCA CTCGACAATA TCGGCCACCA ACAGAAAAA AATCACAATT
 1401 TGCAACTACA GGTATTCAAG CAAATTTTGA AACTTATACT ATTGATTCAA
 1451 ATGAAAAAAG AAATAAAATT AAAAATGGAC ATTTAAATAT TAGGTAA

This encodes a protein having amino acid sequence [[<SEQ ID 172; ORF29ng-1>](#)] ([SEQ ID NO: 172; ORF29ng-1](#)):

1 MNLPIQKFMM LLAAISMLH IPISHANGLD ARLRDDMQAK HYEPGGKYHL
 51 FGNARGSVKN RYCAVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG
 5 GYPEPQGARD IYSYHIKGTSTTKINTVPQ APFSDRWLKE NAGAASGFLS
 201 RADEAGKLIW ENDPKKNWRA NRMDDIRGIV QGAVNPFLTQ FQGVGIGAIT
 251 DSAVSPVTD TAAQQTLOGIN DLGNLSPEAQ LAAASLLQDS AFAVKDGIN
 301 ARQWADAHN ITATAQTALA VAEAAGTVWR GKKVELNPTK WDWVKNTRYK
 351 KPAARHMQTV DGEMAGGNRP PKSITSEGKA NAATYPKLVN QLNEQNLNNI
 401 AAQDPRLSLA IHEGKNFPI GTATYEEADR LGKIWVGEGA RQTSGGGWLS
 10 451 RDGTRQYRPP TEKKSQFATT GIQANFETYT IDSNEKRNKI KNGHLNIR*

ORF29ng-1 (SEQ ID NO: 172) and ORF29-1 (SEQ ID NO: 166) show 86.0% identity in 401 aa overlap:

15	orf29ng-1.pep	MNLPIQKFMM	LLAAISMLH	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL	FGNARGSVKN
	orf29-1	MNLPIQKFMM	LLAAISMLH	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL	FGNARGSVKN
20	orf29ng-1.pep	RVCAYQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG	HEVHSPFDNH	DSKSTSDFSG
	orf29-1	RVCAYQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG	HEVHSPFDNH	DSKSTSDFSG
25	orf29ng-1.pep	GVDGGFTVYQ	LHRTGSEIHP	ADGYDGPQGG	GYPEPQGARD	IYSYHIKGTST	TKTKINTVPQ
	orf29-1	GVDGGFTVYQ	LHRTGSEIHP	ADGYDGPQGG	GYPEPQGARD	IYSYHIKGTST	TKTKINTVPQ
30	orf29ng-1.pep	APFSDRWLKE	NAGAASGFLS	RADEAGKLIW	ENDPKKNWRA	NRMDDIRGIVQ	QAVNPFLTQ
	orf29-1	APFSDRWLKE	NAGAASGFLS	RADEAGKLIW	ENDPKKNWRA	NRMDDIRGIVQ	QAVNPFLTQ
35	orf29ng-1.pep	FQGVGIGAIT	DSAVSPVTD	TAAQQTLOGI	NLGNLSPEAQ	LAAASLLQDS	SAFAVKDGIN
	orf29-1	FQGVGIGAIT	DSAVSPVTD	TAAQQTLOGI	NLGNLSPEAQ	LAAASLLQDS	SAFAVKDGIN
40	orf29ng-1.pep	ARQWADAHN	ITATAQTALA	VAEAAGTVWR	GKKVELNPTK	WDWVKNTRYK	KPAARHMQTV
	orf29-1	ARQWADAHN	ITATAQTALA	VAEAAGTVWR	GKKVELNPTK	WDWVKNTRYK	KPAARHMQTV
45	orf29ng-1.pep	DGEMAGGNRP	PKSITSEGKA	NAATYPKLVN	QLNEQNLNNI	AAQDPRLSLA	IHEGKNFPI
	orf29-1	DGEMAGGNRP	PKSITSEGKA	NAATYPKLVN	QLNEQNLNNI	AAQDPRLSLA	IHEGKNFPI
50	orf29ng-1.pep	IGTATYEEADR	LGKIWVGEGA	RQTSGGGWLS	RDGTRQYRPP	TEKKSQFATT	GIQANFETYT
	orf29-1	IGTATYEEADR	LGKIWVGEGA	RQTSGGGWLS	RDGTRQYRPP	TEKKSQFATT	GIQANFETYT

orf29-1 RYTSLDGKITIIKDNENNYFRIHDNSRKQYLD SNGNAVKTGNLQGKQAKDY LQQQTHIRN
 430 440 450 460 470 480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 21

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 173>] (SEQ ID NO: 173):

10 1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTGAAGAC CAAATGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAG...

This corresponds to the amino acid sequence [<SEQ ID 174; ORF30>] (SEQ ID NO: 174; ORF30):

1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence [<SEQ ID 175>] (SEQ ID NO: 175):

20 1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTGAAGAC CAAAGTGTTC
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 151 ATGAAGGAGA CAGAGGGGGC GTTCTCTCCA TTGGCTATCT TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 25 301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA
 401 GAACAGGTCA TCCTATTGGA AAATTTCCTC ATTATCATCG TCGAGTTACG
 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAAACCGC TTCTAA

This corresponds to the amino acid sequence [<SEQ ID 176; ORF30-1>] (SEQ ID NO: 176; ORF30-1):

35 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKETEGAFLE LAILGGAAIG MWTQHGFSYA TTGRPASVRD VAIAGGLGAI
 101 PGGVGAGKVV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
 151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF30 (SEQ ID NO: 174) shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) (SEQ ID NO: 178) from strain A of *N. meningitidis*:

```

      10      20      30      40
orf30.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
5          |||||
orf30a     MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
          10      20      30      40      50      60

orf30a     LXILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI
          70      80      90     100     110     120

```

The complete length ORF30a nucleotide sequence [<SEQ ID 177>] (SEQ ID NO: 177) is:

```

      1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
     51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
    101  ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
    151  ATGAAGGANA CAGNNGGGGC GTTCTCTCCA TTGGNTATCT TGGGTGGTGC
    201  TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
    251  GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
    301  CCTGGTGNTG TAGGCCCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
    351  ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA
    401  GAACAGGTCA TCCTATTGGN AAATTTCCTT ATTATCATCG TCGAGTTACG
    451  GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
    501  TTGGGAATCA AAATCTACGG ACAGATCATG GAAAAACCGC TTCTAA

```

This encodes a protein having amino acid sequence [<SEQ ID 178>] (SEQ ID NO: 178):

```

    25      1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
          51  MKXTXGAFLP LXILGGAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
    101  PGXVGAAGKV VSFAKYGREI KIGNNMRIAP FGNRTGHPIG KPPHYHRRVT
    151  DNTGKTLPGQ GIGRHRPWES KSTRSWKNR F*

```

ORF30a (SEQ ID NO: 178) and ORF30-1 (SEQ ID NO: 176) show 97.8% identity in 181 aa overlap:

```

orf30a.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60
          |||||
orf30-1     MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP 60

orf30a.pep  LXILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI 120
          | |||||
orf30-1     LAILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI 120

orf30a.pep  KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTRSWKNR 180
          |||||
orf30-1     KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTRSWKNR 180

orf30a.pep  FX
          ||
orf30-1     FX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 (SEQ ID NO: 174) shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) (SEQ ID NO: 180) from *N. gonorrhoeae*:

```

      orf30.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ      42
5      orf30ng      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP      60

```

The complete length ORF30ng nucleotide sequence [<SEQ ID 179>] (SEQ ID NO: 179) is

```

10      1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATCGCCCC
      51  CGCAATGGCA AACCGATTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
      101 ACACGCGGGC AGATGCGCCG ATGCAGTTGG CGGAGCTTTC TCAGAAGGAG
      151 ATGAAGGAGA CTGAAGGGGC TTTTCTTCCA TTGGCTATCT TGGGTGGTGC
      201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
      251 GACCAGCTTC TGTTAGAGAT GTTGCTGGCG GATTAGGCGC AATTCCTGGT
      301 GATGTAGGTG CTGCAGGAAA GGTGTTTCC TTTGCTAAAT ATGGACGTGA
      351 GATTAAATC GGCAATAATA TGCGGATAGC CCCTTTCGGT AATAGAACAG
      401 GTCATCCTAT TGGAAAATTT CCCCATTATC ATCGTCGAGT TACGATAAT
      451 ACGGGCAAGA CTTTGCCTGG ACAGGGAATT GGTCGTCATC GCCCTTGGGA
      501 ATCAAAATCT ACGGACAGAT CATGGAAAAA CCGCTTCTAA

```

20 This encodes a protein having amino acid sequence [<SEQ ID 180>] (SEQ ID NO: 180):

```

25      1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
      51  MKETEGAFLP LAILGGAAIG MWTQHGFSA TTGRPASVRD VAGGLGAIPG
      101 DVGAAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN
      151 TGKTLPGQGI GRHRPWESKS TDRSWKNRF*

```

ORF30ng (SEQ ID NO: 180) and ORF30-1 (SEQ ID NO: 176) show 98.3% identity in 181 aa overlap:

```

30      10      20      30      40      50      60
      orf30ng.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
      orf30-1          MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP

35      70      80      90      100     110
      orf30ng.pep      LAILGGAAIGMWTQHGFSAATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI
      orf30-1          LAILGGAAIGMWTQHGFSAATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI

40      120     130     140     150     160     170
      orf30ng.pep      KIGNNMRIAPFGNRTGHPIGKPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
      orf30-1          KIGNNMRIAPFGNRTGHPIGKPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR

45      180
      orf30ng.pep      FX
      orf30-1          FX

```


Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 22

- 5 The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 181] (SEQ ID NO: 181):

```

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
51 GrTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151 GCACCTGTTT GTg.CGTtAc AAATATCTTT TCTTTTCTT TATTGGGCTT
201 TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA
251 TT..

```

- 15 This corresponds to the amino acid sequence [SEQ ID 182; ORF31] (SEQ ID NO: 182; ORF31):

```

1 MNKTLYRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVFPFGTTH
51 APVCXVTNIF SPSLLGFSLC LAVGTXNIAF ADGI..

```

Further work revealed a further partial nucleotide sequence [SEQ ID 183] (SEQ ID NO: 183):

```

20 1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
51 GGTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151 GCACCTGTTT GTCGTTCAAA TATCTTTTCT TTTTCTTTAT TGGGCTTTTC
25 201 TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGCT GATGGCATT..

```

This corresponds to the amino acid sequence [SEQ ID 184; ORF31-1] (SEQ ID NO: 184; ORF31-1):

```

1 MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPFGTTH
51 APVCRSNIFS FSLGFSLLCL AVGTANIAFA DGI..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 (SEQ ID NO: 182) shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) (SEQ ID NO: 186) from *N. gonorrhoeae*:

orf31ng	 MNKTLYRVIFNKRKGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAF	54
orf31.pep	SFSLLGFSLCLAVGTXNIAFADGI	84
orf31ng	 CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSV	114

The complete length ORF31ng nucleotide sequence [<SEQ ID 185>] (SEQ ID NO: 185) is:

	1	ATGAACAAAA	CCCTCTATCG	TGTGATTTTC	AACCGCAAAAC	GCGGTGCTGT
10	51	GGTAGCTGTT	GCCGAAACCA	CCAAGCGCGA	AGGTAAAAGC	TGTGCCGATA
	101	GTGGTTCGGG	CAGCGTTTAT	GTGAAATCCG	TTTCTTTTCAT	TCCTACTCAT
	151	TCCAAAGCCT	TTTGTTTTTC	TGCATTAGGC	TTTTCTTTAT	GTTTGGCTTT
	201	GGGTACGGTC	AATATTGCTT	TTGCTGACGG	CATTATTACT	GATAAAGCTG
	251	CTCCTAAAAC	CCAACAAGCC	ACGATTCTGC	AAACAGGTaa	cGGCATACCG
15	301	CAAGTCAATA	TTCAAAACCC	TACTTCGGCA	GGGGTTTCTG	TTAATCAATA
	351	TGCCCAAGTTT	GATGTGGGTA	ATCGCGGGGC	GATTTTAAAC	AACAGTCGCA
	401	GCAACACCCA	AACACAGCTA	GGCGGTTGGA	TTCAAGGCAA	TCCTTGGGTG
	451	ACAAGGGGCG	AAGCACGTGT	GGTTGTAAAC	CAATCAACA	GCAGCCATCC
	501	TTCACAACTG	AATGGCTATA	TTGAAGTGGG	TGGACGACGT	GCAGAAGTCG
20	551	TTATTGCCAA	TCCGGCAGGG	ATTGCAGTCA	ATGGTGGTGG	TTTTATCAAT
	601	GCTTCCCGTG	CCACTTTGAC	GACAGGCCAA	CCGCAATATC	AAGCAGGAGA
	651	CTTTAGCGGC	TTTAAGATAA	GGCAAGGCAA	TGCTGTAATC	CGCGGACACG
	701	GTTTGGATGC	CCGTGATACC	GATTTACAC	GTATTCTTGT	ATGCCAACAA
	751	AATCACCTTG	ATCAGTACGG	CCGAACAAGC	AGGCATTTCGT	AA

25 This encodes a protein having amino acid sequence [<SEQ ID 186>] (SEQ ID NO: 186):

30

1	MNKTLYRVIF	NRKRGAVVAV	AETTKREGKS	CADSGSGSVY	VKSVSFIPTH
51	SKAFCFSALG	FSLCLALGTV	NIAFADGIIT	DKAAPKTQQA	TILQTGNIGIP
101	QVNIQTPTSA	GVSVNQYAQF	DVGNRGAILN	NSRNTQTQL	GGWTQGNPWL
151	TRGEARVVVN	QINSSHPSQL	NGYIEVGRR	AEVVIANPAG	IAVNGGGFIN
201	ASRATLTGTQ	PQYQAGDFSG	FKIRQGNVAV	AGHGLDARDT	DFTRILVCQQ
251	NHLDQYGRTS	RHS*			

This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein (SEQ ID NO: 1125) from *Erwinia chrysanthemi* (accession number L39897):

orf31ng	96	GNGIPQVNIQTPTSAGVSVNQYAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE	154
		GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L	
HecA	45	GNGVPVVNIATPDASGLSHNRYHDFNVDNRGLILNNGTARLTPSQLGGLIQNNPNLNGRA	104
Orf31ng	155	ARVVVNQINSSSHPSQLNGYIEVGRRAEVVIANPAGIAVNGGGFINASRATLTGTGQPQYQ	214
		A ++N++ S + S+L GY+EV G+ A VV+ANP GI +G GF+N R TLTTG PQ+	
HecA	105	AAAILNEVSPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTPQFD	164
Orf31ng	215	-AGDFSGFKIRQGNNAVIAGHGLDARDTDF	242
		AG SG +R G+ +I G GLDA +D+	
HecA	165	AAGGLSGLDVRGGDILIDGAGLDASRSY	193

Furthermore, ORF31ng (SEQ ID NO: 186) and ORF31-1 (SEQ ID NO: 184) show 79.5% identity in 83 aa overlap:

		10	20	30	40	50	60
	orf31-1.pep	MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS					
5	orf31ng	MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSDSGSVYVKSVSFIPTH----	10	20	30	40	50
		70	80				
	orf31-1.pep	FSLLGFSCLCLAVGTANIAFADGI					
10	orf31ng	FSLGFSCLCLALGTVNIAFADGIIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSVN	60	70	80	90	100
							110

On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

15 **Example 23**

The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 187](#)] ([SEQ ID NO: 187](#)):

	1	ATGAATACTC	CTCCTTTTGT	CTGTTGGATT	TTTTGCAAGG	TCATCGACAA
	51	TTTCGGCGAC	ATCGGCGTTT	CGTGGCGGCT	CGCCCGTGTT	TTGCACCGCG
20	101	AACTCGGTTG	GCAGGTGCAT	TTGTGGACGG	ACGATGTGTC	CGCCTTGCGT
	151	GCGCTTTGCC	CTGATTTGCC	CGATGTTCCC	TGCGTTCATC	AGGATATTCA
	201	TGTCCGCACT	TGGCATTCCG	ATGCGGCAGA	TATTGATACC	GCG..

This corresponds to the amino acid sequence [[SEQ ID 188](#); [ORF32](#)] ([SEQ ID NO: 188](#); [ORF32](#)):

	1	MNTPPFVCWI	FCKVIDNFGD	IGVSWRLARV	LHRELGWQVH	LWTDDVSALR
	51	ALCPDLPDVP	CVHQDIHVRT	WHSDAADIDT	A..	

Further work revealed the complete nucleotide sequence [[SEQ ID 189](#)] ([SEQ ID NO: 189](#)):

30	1	ATGAATACTC	CTCCTTTTGT	CTGTTGGATT	TTTTGCAAGG	TCATCGACAA
	51	TTTCGGCGAC	ATCGGCGTTT	CGTGGCGGCT	CGCCCGTGTT	TTGCACCGCG
	101	AACTCGGTTG	GCAGGTGCAT	TTGTGGACGG	ACGATGTGTC	CGCCTTGCGT
	151	GCGCTTTGCC	CTGATTTGCC	CGATGTTCCC	TGCGTTCATC	AGGATATTCA
	201	TGTCCGCACT	TGGCATTCCG	ATGCGGCAGA	TATTGATACC	GCGCCTGTTC
35	251	CCGATGTGCT	CATCGAAACT	TTGCCTGCG	ACCTGCCCGA	AAATGTGCTG
	301	CACATTATCC	GCCGACACAA	GCCGCTTTGG	CTGAATTGGG	AATATTTGAG
	351	CGCGGAGGAA	AGCAATGAAA	GGCTGCATCT	GATGCCTTCG	CCGCAGGAGG
	401	GTGTTCAAAA	ATATTTTGG	TTTATGGGTT	TCAGCGAAAA	AAGCGGCGGG
	451	TTGATACGCG	AACGTGATTA	CTGCGAAGCC	GTCCGTTTCG	ATACTGAAGC
40	501	CCTGCGAGAG	CGGCTGATGC	TGCCCCGAAA	AAACGCCTCC	GAATGGCTGC
	551	TTTTCGGCTA	TCGGAGCGAT	GTTTGGGCAA	AGTGGCTGGA	AATGTGGCGA
	601	CAGGCAGGCA	GCCCGATGAC	ACTGTTGCTG	GCGGGGACGC	AAATCATCGA
	651	CAGCCTCAAA	CAAAGCGGCG	TTATTCCGCA	AGATGCCCTG	CAAAACGACG
	701	GCGATGTTTT	TCAGACGGCA	TCCGTCCGCC	TCGTCAAAAT	CCCTTTCGTG
45	751	CCGCAACAGG	ACTTCGACCA	ACTGCTGCAC	CTTGCCGACT	GCGCCGTTCAT

5

801	CCGCGGCGAA	GACAGTTTCG	TGCGCGCCCA	GCTTGC GGCG	AAACCC TTCT
851	TTTGGCACAT	CTACCCGCAA	GACGAGAATG	TCCATCTCGA	CAAAC TCCAC
901	GCCTTTTGGG	ATAAGGCACA	CGGTTTCTAC	ACGCCCCGAA	CCGTGT CGCG
951	ACACCGCCGT	CTTTCCGGACG	ACCTCAACCG	CGGAGAGGCT	TTATCCGCAA
1001	CACAACGCCT	CGAATGTTGG	CAAACCTCGG	AACAACATCA	AAACCGCTGG
1051	CGGCAAGGCG	CGGAGGATTG	GAGCCGTTAT	CTTTTCGGGC	AGCCGTCAGC
1101	TCCTGAAAAA	CTCGCTGCCT	TTGTTTCAAA	GCATCAAAAA	ATACGCTAG

This corresponds to the amino acid sequence [[SEQ ID 190; ORF32-1](#)] ([SEQ ID NO: 190](#);

10 ORF32-1):

```

1      MNTPPFVCWI  FCKVIDNFGD  IGVSWRLARV  LHRELGWQVH  LWTDDVSALR
51     ALCPDLPDVP  CVHQDIHVRT  WHSDAADIDT  APVPDVVIET  FACDLPENVL
101    HIIRRHKPLW  LNWEYLSAEE  SNERLHLMP5  PQEGVQKYFW  FMGFSEKSGG
151    LIRERDYCEA  VRFDTALRE  RLMLPEKNAS  EWLLFGYRSD  VWAKWLEMWR
201    QAGSPMTLLL  AGTQIIDSLK  QSGVIPQDAL  QNDGDVFQTA  SVRLVKIPIFV
251    PQQDFDQLLH  LADCAVIRGE  DSFVRAQLAG  KPFFWHIYPQ  DENVHLDKLH
301    AFWDKAHGFP  TPETVSAHRR  LSDDLNGGEA  LSATQRLECW  QTLQQHQNGW
351    RQGAEDWSRY  LFGQPSAPEK  LAAFVSKHQK  IR*w

```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 (SEQ ID NO: 188) shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) (SEQ ID NO: 192) from strain A of *N. meningitidis*:

```

25      orf32.pep      MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
      |||||          ||||||||||||||||||||||||||||||||||||||||||||
      orf32a          MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
                        10          20          30          40          50          60

30      orf32.pep      CVHQDIHVRTWHSDAADIDTA
      |||||          |||||||
      orf32a          CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
                        70          80          90          100         110         120

```

35 The complete length ORF32a nucleotide sequence [<SEQ ID 191>] (SEQ ID NO: 191) is:

	1	ATGAATACTC	CTCCTTTTTC	TGCTGGANTT	TTTTGCAAGG	TCATCGACAA
	51	TTTCGGCGAC	ATCGGCGTTT	CGTGGCGGCT	TGCCCGTGTT	TTGCACCGCG
	101	AACTCGGTTG	GCAGGTGCAT	TTGTGGACGG	ACGATGTGTC	CGCCTTGCGT
40	151	GCGCTTTGCC	CTGATTTGCC	CGATGTTTNC	TGCGTTTCAT	AGGATATTCA
	201	TGTCGCCACT	TGGCATTCCG	ATGCGGCAGA	TATTGATACC	GCGCTGTGTC
	251	NCGATGTCGT	CATCGAAACT	TTTGCCCTGC	ACCTGCCCGA	AAATGTGCTG
	301	CACATCATCC	GCCGACACAA	GCCGCTTTGG	CTGAANTGGG	AATATTTGAG
	351	CGCGGAGGAN	AGCAATGAAA	GGCTGCACNT	GATGCCTTCG	CCGCAGGAGA
45	401	GTGTTCNAAA	ATANTTTTGG	TTTATGGGTT	TCAGCGAANN	NAGCGGCGGA
	451	CTGATACGCG	AACGCGATTA	CTGCGAAGCC	GTCCGTTTCG	ATAGCGGAGC
	501	TTTGCGCAAG	AGGCTGATGC	TTCCCGAAAA	AAACGCCCCC	GAATGGCTGC
	551	TTTTCGGCTA	TCGGAGCGAT	GTTTGGGCAA	AGTGGCTGGA	AATGTGGCGA

5
10
1101

601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNNGGGCGC ANATTATCGA
651 CAGCCTCAAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TCGCGGCCCA GCTTGCGGGC AAACCCTTCT
851 TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
951 ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAATCCTGC AACAACATCA AAACGGCTGG
1051 CGGCAAGGCG CGGAGGATG GAGCCGTAT CTTTTTGGGC AGCCTTCCGC
1101 ATCCGAAAAA CTCGCCGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG

This encodes a protein having amino acid sequence [<SEQ ID 192>] (SEQ ID NO: 192):

15
20
351

1 MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51 ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL
101 HIIRRHKPLW LXWEYLSAEX SNERLHXMPs PQESVXKXFW FMGFSEXSGG
151 LIRERDYCEA VRFDSGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR
201 QAGSPLTLLL AGAXIIDSLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDKLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301 AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLECW QILQQHQNGW
351 RQGAEDWSRY LFGQPSASEK LAAFVSKHOK IR*

ORF32a (SEQ ID NO: 192) and ORF32-1 (SEQ ID NO: 190) show 93.2% identity in 382 aa overlap:

25
30
35
40
45
50

orf32-1.pep	10	20	30	40	50	60
orf32a	10	20	30	40	50	60
orf32-1.pep	70	80	90	100	110	120
orf32a	70	80	90	100	110	120
orf32-1.pep	130	140	150	160	170	180
orf32a	130	140	150	160	170	180
orf32-1.pep	190	200	210	220	230	240
orf32a	190	200	210	220	230	240
orf32-1.pep	250	260	270	280	290	300
orf32a	250	260	270	280	290	300
	310	320	330	340	350	360

```

orf32-1.pep  AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLCQWTLQQHQNGWRQGAEDWSRY
              |||||:|||||
orf32a       AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATQRLCQWILQQHQNGWRQGAEDWSRY
              310      320      330      340      350      360

```

```

5
              370      380
orf32-1.pep  LFGQPSAPEKLAAFVSKHQKIRX
              ||||| |||||
orf32a       LFGQPSAPEKLAAFVSKHQKIRX
              370      380

```

10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 (SEQ ID NO: 188) shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) (SEQ ID NO: 194) from *N. gonorrhoeae*:

```

orf32.pep      MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP 57
              ||| |
15 orf32ng      MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP 60

orf32.pep      DVPCVHQDIHVRTWHSDAADIDTA 81
              ||| |||||
orf32ng        DVPFVHQDIHVRTWHSDAADIDTAPVDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120

```

20 An ORF32ng nucleotide sequence [<SEQ ID 193>] (SEQ ID NO: 193) was predicted to encode a protein having amino acid sequence [<SEQ ID 194>] (SEQ ID NO: 194):

```

      1  MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDDVS
      51  ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVPDAV IETFACDLPE
25 101  NVLNIIRRHK PLWLNWEYLS AESNERLHL MPSPQEGVQK YFWFMGFSEK
      151  SGGIRERDY REAVRFDTEA LRRRLVLP EK NAPEWLLFGY RGDVWAKWLD
      201  MWQQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
      251  PFVPQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFFWHI YPQDENVHLD
      301  KLHAFWDKAY GFYTPETASV HRLLSDDLNG GEALSATQRL ECGVL*

```

30 Further sequencing revealed the following DNA sequence [<SEQ ID 195>] (SEQ ID NO: 195):

```

      1  ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTGTGCA AGGTCATCGA
      51  CAATTTTCGGC GACATCGGCG TTTCTGTTGG GCTCGCCCGT GTTTTGCACC
35 101  GCGAACTCGG TTGGCAGGTG CATTGTGTTG CCGACGACGT GTCCGCCTTG
      151  CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTT ATCAGGATAT
      201  TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
      251  TTCCCGATGC CGTTATCGAA ACTTTTGCTT GCGACCTGCC CGAAAATGTG
      301  CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGGAAATATT
      351  GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGACAG
40 401  AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
      451  GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATACCGA
      501  AGCCCTGCGC CGGCGGCTGG TGCTGCCCGA AAAAAACGCC CCCGAATGGC
      551  TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
      601  CAACAGGCAG GCAGCCTGAT GACCCACTGT CTGGCGGGGG CGCAAATTAT
      651  cagacgcttc AAACAAGCG GCGTTATTCC GCAAAACGCC CTGCAAAAtg
45 701  aaggcggtgt CTTTCagacG gcatccgTcC gccttGTCAA AAtcCCGTTC
      751  GTGCcGCAAC AGGAcTTcGA CAAATGCTG CAcctcgcCG ACTGCGCCGT
      801  GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCTT

```



```

orf32ng-1      HAFWDKAYGYFTPETASVHRLLSDDLNGGEALSATQRLECWQTLQQHQNGWRQGAEDWSR
                310          320          330          340          350          360

orf32-1.pep    360          370          380
                YLFGQPSAPEKLAAFVSKHQKIRX
                ||||| ||||| ||||| ||||| |||||
orf32ng-1      YLFGQPSASEKLAAFVSKHQKIRX
                370          380

```

On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 ([SEQ ID NO: 190](#)) (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 ([SEQ ID NO: 190](#)) is a surface-exposed protein, and that it is a useful immunogen.

Example 24

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 197>] (SEQ ID NO: 197):

```

1      .TTGTTCCTGC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCACGTG
51     GTTTCGGGNC AAAGACCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG
101    ACGAGTGGCG GCA .ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
151    AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGTGTGCGG TATTGTTGCT
201    GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA
251    ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTCGAAA
301    CTCGGTTTC CTGTCCCCCG TGCGCGGTGT GTCATCGAAG GCCGCTGAA
351    CCGCAATATT CCGCATGCGC TGGCTTGGTC GGGGCTGCTG GTCGNAGTA
401    TCGCTGTGTA NGGCATCCTG CCGCGCCTG.

```

This corresponds to the amino acid sequence [<SEQ ID 198; ORF33>] (SEQ ID NO: 198; ORF33):

```

1      .LFLRVKVGRF FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
51     SLWLCTLLGM LVSVLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK
101    LGFPVPDARS VIEGRNLNGNI ADARAWSGLL VXSIACXGIL PRL..

```


Further work revealed the complete nucleotide sequence [<SEQ ID 199>] (SEQ ID NO: 199):

```

1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
51 AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GCGCGAGATG
5  151 ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TCGCTGCGGG
201 GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCATT TTTACCGGTT
251 TTTCAGTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGG
301 GTTTTGGCGG GCGTGTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
10 401 CGACGTGGTT TCGGGGCAAA GACCCTGTAA ATCAGGCGGT GTTGGCGCTG
451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGAAAA TAGGCGCAAC
501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCCGTAT
551 TGTTGCTGCT TTTGGTGCGG CAATATACGT TCAACTGGGA AAGCACGCTG
601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
15 651 GTCGAAACTC GGTTCCTCTG TCCCGATGC GCGGGCGGTC ATCGAAGGCC
701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCCGG GCTGCTGGTC
751 GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTGCTGG CTTGGGTAGT
801 GTGTAAATC CTTTTGAAAA CAAGCGAAAA CGGATTGGAT TTGAAAAGC
851 CCTATTATCA GCGGTCATC CGCCGTGGC AGAACAAAAT CACCGATGCG
20 901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAAA TCATCTTGAA
951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GCGCAACTG CTTATCGGCG TCGCGCCCA AACTGTGCCG GACCGCGCG
25 1151 TGTTGCGGCA GATTGTCCGA CTCTCGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCAGCTTT TGGCGGAACA GGGGCTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
1301 CTGACAGGGC GGCAGGAA GGGCGTTGA AAGACCAATA A

```

30 This corresponds to the amino acid sequence [<SEQ ID 200; ORF33-1>] (SEQ ID NO: 200; ORF33-1):

```

1  MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
51 IDNRNMLRET LERVAGSFW LWVVAATFAF FTGFSVTYLL MDNQGLNFFL
101 VLAGVLGMNT LMLAVWLAML FLRVKVRFF SSPATWFRGK DPNQAVLRL
35 151 YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRLNGNIA DARAWSGLLV
251 GSIACYGILP RLLAWVVKI LLKTSENGLD LEKPYQAVI RRWQNKITDA
301 DTRRETSAV SPKII LNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351 TNREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRIVR LSEAAQGGAV
40 401 VQLLAEQGLS DDLSEKLEHW RNALAECGAA WLEPDRAAQE GRLKDQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 (SEQ ID NO: 198) shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) (SEQ ID NO: 202) from strain A of *N. meningitidis*:

```

orf33.pep
10 20 30
LFLRVKVRFFSSPATWFRXKDPVNQAVLR
|||||

```

5	orf33a	LMDNQGLNFFLVLAGVXGMNTLM LAVWLAMLFLRVKVGRRFFSSPATWFRGKDPVNQAVLR	90	100	110	120	130	140
	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA	40	50	60	70	80	90
10	orf33a	LYADEWRXPSVRWKIGATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLGDSSSVRL	150	160	170	180	190	200
	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGRNLNGNIADARAWSGLLVXSIACXGILPRL	100	110	120	130	140	
15	orf33a	VEMLAWLPAKLGFVPDARAVIEGRNLNGNIADARAWSGLLVGSIACYGILPRLLA WAVCK	210	220	230	240	250	260
	orf33a	ILXXTSENGLDLEKXXXXXXIRRWQNKITDADTRRET VSAVSPKIVLNDAPKWAVMLETE	270	280	290	300	310	320

The complete length ORF33a nucleotide sequence [<SEQ ID 201>] (SEQ ID NO: 201) is:

	1	ATGTTGAATC	CATCCCGAAA	ACTGGTTGAG	CTGGTCCGTA	TTTGGGAAGA
	51	AGGCGGCTTT	ATTTTCAGCG	GCGATCCCGT	GCAGGCGACG	GAGGCTTTGC
20	101	GCCGCGTGGA	CGGCAGTACG	GAGGAAAAAA	TCATCCGTCG	GGCGAAGATG
	151	ATCGACAGGA	ACCGTATGCT	GCGGGAGACG	TTGGAACGTG	TGCGTGCGGG
	201	GTCGTTCTGG	TTGTGGGTGG	CGGCGGCGAC	GTTTGCCTTT	NTTACCGNTT
	251	TTTCAGTTAC	TTATCTTCTA	ATGGACAATC	AGGGTCTGAA	TTCTTTTTTG
	301	GTTTTGGCGG	GCGTGNTGGG	CATGAATACG	CTGATGCTGG	CAGTATGGTT
	351	GGCAATGTTG	TTCCTGCGCG	TGAAAGTGGG	GCGTTTTTTC	AGCAGTCCGG
25	401	CGACGTGGTT	TCGGGGCAAA	GACCTGTCA	ATCAGGCGGT	GTTGCGGCTG
	451	TATGCGGACG	AGTGGCGGCN	ACCTTCGGTA	CGTTGGAAAA	TAGGCGCAAC
	501	GTCGCACAGC	CTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTCGGTAT
	551	TGTTGCTGCT	TTTGGTGGG	CAATATACGT	TCAACTGGGA	AAGCACGCTG
	601	TTGGGCGATT	CGTCTTCGGT	ACGGCTGGTG	GAAATGTTGG	CATGGCTGCC
30	651	TGCGAAACTG	GGTTTTCCCG	TGCCTGATGC	GCGGGCGGTC	ATCGAAGGTC
	701	GTCTGAACGG	CAATATTGCC	GATGCGCGGG	CTTGGTCGGG	GCTGCTGGTC
	751	GGCAGTATCG	CCTGTACGG	CATCCTGCCG	CGCCTCTTGG	CTGGGCGGGT
	801	ATGCAAAATC	CTTNTGNAAA	CAAGCGAAAA	CGGCTTGAT	TTGAAAAAGC
	851	NCNNNNNTCN	NNCGNTCATC	CGCCGCTGGC	AGAACAAAAT	CACCGATGCG
35	901	GATACGCGTC	GGGAAACCGT	GTCCGCGGTT	TCGCCGAAAA	TCGCTTTGAA
	951	CGATGCGCCG	AAATGGGCGG	TCATGCTGGA	GACCGAATGG	CAGGACGGCG
	1001	AATGGTTCGA	GGGCAGGCTG	GCGCAGGAAT	GGCTGGATAA	GGGCGTTGCC
	1051	GCCAATCGGG	AACAGGTTGC	CGCGCTGGAG	ACAGAGCTGA	AGCAGAAACC
	1101	GGCGCAACTG	CTTATCGGCG	TGCGCGCCCA	AACTGTGCC	GACCGCGGCG
40	1151	TGTTGCGGCA	GATCGTCCGA	CTTTCGGAAG	CGGCGCAGGG	CGGCGCGGTC
	1201	GTGCANCTTT	TGGCGGAACA	GGGGCTTTCA	GACGACCTTT	CGGAAAAGCT
	1251	GGAACATTGG	CGTAACGCGC	TGACCGAATG	CGGCGCGGCG	TGGCTGGAAC
	1301	CCGACAGAGC	GGCGCAGGAA	GGCCGTCTGA	AAACCAACGA	CCGCACTTGA

45 This encodes a protein having amino acid sequence [<SEQ ID 202>] (SEQ ID NO: 202):

	1	MLNPSRKLVE	LVRILEEGGF	IFSGDPVQAT	EALRRVDGST	EEKIIRRAKM
	51	IDRNRMLRET	LERVAGSFW	LWVAAATFAF	XTXFSVTYLL	MDNQGLNFFL
	101	VLAVXGMNT	LMLAVWLAML	FLRVKVGRRF	SSPATWFRGK	DPVNQAVLRL
	151	YADEWRXPSV	RWKIGATSHS	LWLCTLLGML	VSVLLLLLV	QYTFNWESTL
50	201	LGDSSSVRLV	EMLAWLPAKL	GFPVPDARAV	IEGRNLNGNIA	DARAWSGLLV
	251	GSIAICYGILP	RLLAWAVCKI	LXXTSENGLD	LEKXXXXXXI	RRWQNKITDA
	301	DTRRET VSAV	SPKIVLNDAP	KWAVMLETEW	QDGEWFEGRL	AQEWLDKGVA
	351	ANREQVAALE	TELKQKPAQL	LIGVRAQTVP	DRGVLRQIVR	LSEAAQGGA

401 VXLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRLKTNDRT*

ORF33a (SEQ ID NO: 202) and ORF33-1 (SEQ ID NO: 200) show 94.1% identity in 444 aa overlap:

```

5      10      20      30      40      50      60
orf33a.pep  MLNPSRKLVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAKMIDNRMLRET
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf33-1      MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRMLRET
           10      20      30      40      50      60

10     70      80      90      100     110     120
orf33a.pep  LERVVAGSFWLWVAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf33-1      LERVVAGSFWLWVAATFAFFTGFSTYLLMDNQGLNFFLVLAGVLMNTLMLAVWLAML
           70      80      90      100     110     120

15     130     140     150     160     170     180
orf33a.pep  FLRVKVGRFFSSPATWFRGKDPVNQAVLRRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf33-1      FLRVKVGRFFSSPATWFRGKDPVNQAVLRRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
           130     140     150     160     170     180

20     190     200     210     220     230     240
orf33a.pep  VSVLLLLLVQRQYTFNWESTLLGDSSSVRLVEMLAWLPAKLGFPVPDARAVIEGRNLGNIA
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf33-1      VSVLLLLLVQRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
           190     200     210     220     230     240

25     250     260     270     280     290     300
orf33a.pep  DARAWSGLLVGSIACYGILPRLLAWAVCKILXXTSENGLDLEKXXXXXXIRRWNQKITDA
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf33-1      DARAWSGLLVGSIACYGILPRLLAWVCKILLKTSSENGLDLEKPYQAVIRRWNQKITDA
           250     260     270     280     290     300

30     310     320     330     340     350     360
orf33a.pep  DTRRETSAVSPKIVLNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVAANREQVAALE
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf33-1      DTRRETSAVSPKIIILNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVATNREQVAALE
           310     320     330     340     350     360

35     370     380     390     400     410     420
orf33a.pep  TELKQKPAQLLIGVRAQTVPDGRVLRQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf33-1      TELKQKPAQLLIGVRAQTVPDGRVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW
           370     380     390     400     410     420

40     430     440     450
orf33a.pep  RNALTECGAAWLEPDRAAQEGRLKTNDRTX
           |||:|||||:|||||:|||||
orf33-1      RNALAECCGAAWLEPDRAAQEGRLKDQX
           430     440

```

45 Homology with a predicted ORF from *N.gonorrhoeae*

1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
 1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC
 1301 CTGACAGGGT GGCACAGGAA GGCCGTTTGA AAGACCAATA A

- 5 This encodes a protein having amino acid sequence [<SEQ ID 206; ORF33ng-1>] (SEQ ID NO: 206; ORF33ng-1):

1 MLNPSRKLVE LVRILNKGGF IFSGDPVQAT EALRRVDGST EEKIFRAEM
 51 IDRDRMLRDT LERVAGSFW LWVVASMMF TAGFSGTYLL MDNQGLNFFL
 101 VLAGVLGMNT LMLAVWLATL FLRVKVGRRF SSPATWFRGK GPVNQAVLRL
 10 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
 201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV
 251 GSIVCYGILP RLLAWVVKI LLKTSENGLD LEKTYQAVI RRWQNKITDA
 301 DTRRETSAV SPKIVLNDAP KWALMLETW QDGQWFEGR LAQEWLDKGVA
 351 ANREQVALE TELKQKPAQL LIGVRAQTP DRGVLQIVR LSEAAQGGAV
 15 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ*

ORF33ng-1 (SEQ ID NO: 206) and ORF33-1 (SEQ ID NO: 200) show 94.6% identity in 446 aa overlap:

20	orf33-1.pep	MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRMLRET
	orf33ng-1	MLNPSRKLVELVRILNKGGFIFSGDPVQATEALRRVDGSTEEKIFRAEMIDRDRMLRDT
25	orf33-1.pep	LERVRAGSFWLWVVAATFAFFTGFSVTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLAML
	orf33ng-1	LERVRAGSFWLWVVVASMMFTAGFSGTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLATL
30	orf33-1.pep	FLRVKVGRRFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
	orf33ng-1	FLRVKVGRRFSSPATWFRGKGPVNQAVLRLYADQWRQPSVRWKIGATAHSWLCTLLGML
35	orf33-1.pep	VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFVPDARAVIEGRNLGNIA
	orf33ng-1	VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFVPDARAVIEGRNLGNIA
40	orf33-1.pep	DARAWSGLLVGSIVCYGILPRLAWVVKI LLKTSENGLDLEKPYQAVIRRWQNKITDA
	orf33ng-1	DARAWSGLLVGSIVCYGILPRLAWVVKI LLKTSENGLDLEKTYQAVIRRWQNKITDA
45	orf33-1.pep	DTRRETSAVSPKII LNDAPKWAVMLETWQDGQWFEGR LAQEWLDKGVA TNREQVALE
	orf33ng-1	DTRRETSAVSPKIVLNDAPKWALMLETWQDGQWFEGR LAQEWLDKGVA ANREQVALE

-206-

		370	380	390	400	410	420
	orf33-1.pep	TELKQKPAQLLIGVRAQTV	PD	RGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW			
5	orf33ng-1	TELKQKPAQLLIGVRAQTV	PD	RGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW			
		370	380	390	400	410	420
		430	440				
	orf33-1.pep	RNALAECGAAWLEPDRAAQEGRLKDQX					
10	orf33ng-1	RNALTECGAAWLEPDRVAQEGRLKDQX					
		430	440				

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 25

- 15 The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 207](#)] ([SEQ ID NO: 207](#)):

20
25
1 ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTCGGCG GCGTGTTCCTT
51 CGGGGTGTCC GGTCTGGTAT GGTTCCTTT GGGCGTTTCT TT.GAGTGCG
101 CCTGTTTTTC GGTGTTTTCT TTTCGGGGTT CGGGACGGGG GACGTTTG
151 GGCAGTACGG GGGTTTCTTT GAGTGTGTT TCAGCTTGTT TTCC.GGCGT
201 CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTTG CG..GTTTGA
251 CCCGTTTTTT CTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
301 TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
351 GGGTTGGGCG GCATCTTGTT CCGACTACGC CGTTTGGCAG CCAGAATTCG
401 GTTTCGCGGG GGTGTGCGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
451 GTCC..

This corresponds to the amino acid sequence [[SEQ ID 208; ORF34](#)] ([SEQ ID NO: 208; ORF34](#)):

30
1 ..QKSLSRISLW GLGGVFFGVS GLVWFSLGVS XECACFSGVS FRGSGRGTFV
51 GSTGVSLSVF SACVXGVVRL PVGLSCVGRL XXLTRFFLGA AGDVILLPLS
101 SVPSGCAGSD EAAWWCAGWA ASCPTTPFGS QNSVSRGLSV CCGSA*RVLS
151 S..

- 35 Further work revealed the complete nucleotide sequence [[SEQ ID 209](#)] ([SEQ ID NO: 209](#)):

40
1 ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
51 GCCGGGTCAG AATAGGTTGT CCAGAATTC TTTATGGGGT TTGGGCGGCG
101 TGTTTTTCGG GGTGTCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
151 GGCTGCGCCT GTTTTCGGG TGTTCCTTT CGGGGTTCCG GACGGGGGAC
201 GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTTCA GCTGTGTTC
251 CGGCGTCGTC CGGCTGCCTG TCGGTTTGAG CTGTGTCGGC AGGTGCGGT
301 TTGACCCGGT TTTTCTTGGG TCGGCGAGG GACGCGAGTC CGTGCCCGT
351 TTCGTCTGTG CCGTCCGGT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT

5	401	GTTCGGGTTG	GGCGGCATCT	TGTCGGAATA	CGCCGTTTGG	CAGCCAGAAAT
	451	TCGGTTTCGC	GGGGGCTGTC	GGTGTGTTGC	GGTTCGGCTT	GAAGGGTTTT
	501	GTCGCCGTTC	GGGTTGAATG	TGCTGACGAT	GCCTATTGCC	AATGCGCCGA
	551	TGGCGGCGAT	ACAGATGAGC	AATACGGCGC	GTATCAGGAG	TTTGGGGGTC
	601	AGCCTGAAGG	GTTTGTTCGG	TTTTTTTGCC	ATTTTGATTG	TGCTTTTGGG
10	651	GTGTCGGGCA	ATGCCGTCTG	AAGGCGGTTG	AGACGGCATT	GCCGAGTCAG
	701	CGTTGGACGT	AGTTTTGGTA	GAGGGTGATG	ACTTTTGTGA	CGCCGACGGT
	751	GGTGCTGACT	TTTTGGGTAA	TCTGCGCCTG	TTCTTCGGGG	GTGAGGATGC
	801	CCATAACGTA	GGTTACGTTG	CCGTAGGTAA	CGATTTTGAC	GC GCGCCTGT
	851	GTGGCGGGGC	TGATGCCCAA	CAGCGTGCGG	CGGACTTTGG	ATGTGTTCCA
15	901	AGTGTCGCCG	GCGATGTCGC	CGGCAGTGCG	CGGCAGGGAG	GCGACGGTAA
	951	TATAGTTGTA	CACGCCCTTC	GCGGCTTGTT	CGGAACGTGC	AATCTGACCG
	1001	ACGAACTGTT	TTTCGCCTTC	GGTGGCGACT	TGTCGAGACA	GCAGCAGGTG
	1051	GCGGTTGTAG	CCGACGACGG	AGATTTGGGG	CGTGTAACCT	TTGGTTTGGT
	1101	TGTTTTGGCG	CAGATAGGAA	CGGGCGGTGG	TTTCGATACG	CAACGCCATA
20	1151	ACGTTGTTCG	CGGTTTGCGC	GCCGGTGGTT	CGGCGGTAGA	CGGCGGATTT
	1201	CGCGCCGACG	CGGCGGCTTC	CGATTACTGC	GCTGACGCGA	CCGCTAAGGG
	1251	CAAGGCTGAA	AATGGCGGCA	ATCAGGGTGC	GGACGGTGTG	CGGTTTGGGT
	1301	TTCATCGGGT	GCTTCCTTTC	TTGGGCGTTT	CAGACGGCAT	TGCTTTGCGC
	1351	CATGCCGTCT	GA			

This corresponds to the amino acid sequence [[<SEQ ID 210; ORF34-1>](#)] ([SEQ ID NO: 210; ORF34-1](#)):

25 1 M M P P F I M L P W I A G V P A V P G Q N R L S R I S L W G L G G V F F G V S G L V W F S L G V S L
 51 G C A C F S G V S F R G S G R G T F V G S T G V S L S V F S A C V P A S S G C L S V * A V S A G C G
 101 L T R F F L G A A G D G S P L P L S S V P S G C A G S D E A A W W C S G W A A S C P T T P F G S Q N
 151 S V S R G L S V C C G S A * R V L S P F G L N V L T M P I A N A P M A A I Q M S N T A R I R S L G V
 201 S L K G L F G F F A I L I V L L G C R A M P S E G G S D G I A E S A L D V V L V E G D D F L Y A D G
 251 G A D F L G N L R L F F G G E D A H N V G Y V A V G N D F D A R L C G G A D A Q Q R G A D F G C V P
 30 301 S V A G D V A G S A R Q G G D G N I V V H A F G G L F G T C N L T D E L F F A F G G D L S E Q Q Q V
 351 A V V A D D G D L G R V A F G L V V L A Q I G T G G G F D T Q R H N V V V G L R A G G S A V D G G F
 401 R A D G G A S D Y C A D A A A K G K A E N G G N Q G A D G V R F G F H R V L P F L G V S D G I A L R
 451 H A V *

35 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF34 (SEQ ID NO: 208) shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) (SEQ ID NO: 212) from strain A of *N. meningitidis*:

[illegible]

```

      100      110      120      130      140      150
orf34.pep  AGDVILLPLSSVPSGCAGSDEAAWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLS
           ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
orf34a     AGDGSPLPLSSVPSGCAGADEEAXXCSGWAASCPTTPFGSQNSVSRGLSVCCGSVWRVLS
           120      130      140      150      160      170

orf34.pep  S
orf34a     PFGXNVLTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSD
           180      190      200      210      220      230

```

The complete length ORF34a nucleotide sequence [<SEQ ID 211>] (SEQ ID NO: 211) is:

```

      1  ATGATGATNC CGTTNATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
     51  GCCGGGTCAG AAGAGGTTGT CGAGAANTTC TTTATGGGGT TTAGGCGGCN
    101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTNNT
    151  TCTTTGGGTG TTTCTNTGGG CTGTGCCTGT TTTTCGGGTG TTTCTTTTCG
    201  GGGTTCGGGA CGGGGACGT TTGTGGGCAG TACNGGGGTT TCTTTGAGTG
    251  TGTTTTCAGC TTGTGCTCCG GCGTCGTCCG GCTGCCTGTC GGTTTNAGCT
    301  GTGTCGGCAG GTTGCGGTTT GACCCGNTT TTCTTNGGTG CGGCAGGGGA
    351  CGGCAGTCCG CTGCCGCTTT CGTCTGTGCC GTCCGGCTGT GCGGGTGC GG
    401  ATGAGGAGGC GTNGTNGTGT TCGGGTTGGG CGGCATCTTG TCCGACTACG
    451  CCGTTTGGCA GCCAGAATTC GGTTCGCGG GGGCTGTCCG TGTGTTGCGG
    501  TTCGGTNTGG AGGGTTTTGT CNCCGTTCCG GTNGAATGTG CTGACGATGC
    551  CTATTGCCAA TGCGCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
    601  ATCAGGAGTT TGGGGGTCAG CCTGAAGGGT TTGTTCNGTT TTTTGTCCAT
    651  TTTGATTGTG CTTTGTGGGT GTCGGGCAAT GCCGTCTGAA GGCGGTTT CAG
    701  ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTNNGGTAGA GGGTGATGAC
    751  TTTTTGTACG CCGACGTGGG TGCTGACTTT TTGGGTAATC TGCGCCTGTT
    801  CTTCCGGGGT GAGGATGCCC ATAACGTAGG TTACGTTGCC GTAGGTAACG
    851  ATTTTGACGC GCGCCTGTGT GCGGGGGCTG ATGCCCAACA GCGTGGCGCG
    901  GACTTTGGAT GTGTTCCAAG TGTCGCGCGC GATGTCGCGG GCAGTGCGCG
    951  GCAGGGAGGC GACGTAATG TANTTGTACA CGCCTTCGGC GGCCTGTTTCG
   1001  GAACGTGCAA TCTGACCGAC GAACTGTTTC TCGCCTTCGG TGGCGACTTG
   1051  TCCGAGCAGC AGCAGGTGGC GGTGTAGACC GACAACGGAG ATTGGGGGCG
   1101  TGTANCCTTT GGTTTGTTTG TTTTGGCGCA GATAGGAGCG GGCCTGTTTCG
   1151  TCGATACGCA GCGCCATTAC GTTGTCTGTC GTTNGCGCGC CGGTGGTTTCG
   1201  GCGGTCGACG GCGGATTTCG CGCCGACCGC CGCGCCGCGC ACGACTGCGC
   1251  TGACGCAGCC GCCGAGGCA AGGCTGAGGA CGGCGGCAGT CAGGTTGCGG
   1301  ACGGTGTGCG GTTTGGGTTT CATCGGTGTC TTCCTTTCTT GGGCGTTTCA
   1351  GACGGCATTG CTTTGCGCCA TGCCGTCTGA

```

This encodes a protein having amino acid sequence [<SEQ ID 212>] (SEQ ID NO: 212):

```

      1  MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLGVSX
     51  SLGVSXGCAC FSGVSFRGSG RGTFVGSTGV SLSVFSACAP ASSGCLSVXA
    101  VSAGCGLTRX FXGAAGDGSP LPLSSVPSGC AGADEEAXXC SGWAASCPTT
    151  PFGSQNSVSR GLSVCCGSVW RVLSPFGXNV LTMPIANAPM AVIQMSNTAR
    201  IRS LGVSLKG LFXFFAILIV LLGCRAMPSE GSGDGLAESA LDVVXVEGDD
    251  FLYADGGADF LGNLR LFFGG EDAHNVGYVA VGNDFDARLC GGADAQQRGA
    301  DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL
    351  SEQQQVAVVA DNGDLGRVXF GLVVL AQIGA GGGFD TQRHY VVGXRAGGS
    401  AVDGGFRADR RAADD CADAA AEGKAEDGGS QGADGVRF GF HRVLPFLGVS
    451  DGIALRHAV*

```

ORF34a (SEQ ID NO: 212) and ORF34-1 (SEQ ID NO: 210) show 91.3% identity in 459 aa overlap:

		10	20	30	40	50	60
	orf34a.pep	MMXPXIMLPW	IAGVPAVPGQ	KRLSRXSLWGL	GGXFFGVSG	LVWVSLGV	SXSLSGVSGXCAC
5	orf34-1	MMMPFIMLPW	IAGVPAVPGQ	NRLSRISLWGL	GGVFFGVSG	LVWVSLGV	SL-----GCAC
		10	20	30	40	50	
		70	80	90	100	110	120
	orf34a.pep	FSGVSRFGS	GRGTFVGST	GVSLSVFS	SACAPASSG	CLSVXAVS	AGCLTRXFXGAAGDGSP
10	orf34-1	FSGVSRFGS	GRGTFVGST	GVSLSVFS	SACVPASSG	CLSVXAVS	AGCLTRFFLGAAGDGSP
		60	70	80	90	100	110
		130	140	150	160	170	180
	orf34a.pep	LPLSSVPSG	CAGADEEAX	XCSGWAAS	CPTTFPGS	QNSVSRGL	SVCCGSVVRVLS
15	orf34-1	LPLSSVPSG	CAGSDEAA	WCSGWAAS	CPTTFPGS	QNSVSRGL	SVCCGSAXRVLS
		120	130	140	150	160	170
		190	200	210	220	230	240
	orf34a.pep	LTMPIANAP	MAVIQMSNT	ARIRSLGV	SLKGLFXF	FAILIVLL	GCRAMPSEGGSDGIAESA
20	orf34-1	LTMPIANAP	MAAIQMSNT	ARIRSLGV	SLKGLFGF	FAILIVLL	GCRAMPSEGGSDGIAESA
		180	190	200	210	220	230
		250	260	270	280	290	300
	orf34a.pep	LDVVXVEGD	DFLYADGG	ADFLGNL	RLLFFGG	EDAHNVGY	VAVGNDFDARL
25	orf34-1	LDVVLVEGD	DFLYADGG	ADFLGNL	RLLFFGG	EDAHNVGY	VAVGNDFDARL
		240	250	260	270	280	290
		310	320	330	340	350	360
	orf34a.pep	DFGCVPSV	AGDVAGS	ARQGGD	GNVXVH	AFGGLFG	TCLNLTDELFL
30	orf34-1	DFGCVPSV	AGDVAGS	ARQGGD	GNIVVH	AFGGLFG	TCLNLTDELFF
		300	310	320	330	340	350
		370	380	390	400	410	420
	orf34a.pep	DNGDLGRV	XFGLVVL	AQIGAGG	GFDTORH	VVVGX	RAGGS
35	orf34-1	DDGDLGRV	AFGLVVL	AQIGTGG	GFDTORH	VVVGL	RAGGS
		360	370	380	390	400	410
		430	440	450	460		
	orf34a.pep	AEGKAEDG	GSQGADG	VRFGFH	RVLFP	FLGVSDG	IALRHAVX
40	orf34-1	AKGKAENG	GNQGADG	VRFGFH	RVLFP	FLGVSDG	IALRHAVX
		420	430	440	450		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF34 (SEQ ID NO: 208) shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) (SEQ ID NO: 214) from *N. gonorrhoeae*:

```

45      orf34 .pep                               QKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE-----CAC      35
      || |||||:|||||||
      orf34ng  MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLGVSFSLGVS LGCAC      60

```

```

orf34.pep    FSGVSFRGSGRGTFTVGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXXLTRFFFLGA    90
|||||      |:|||||      |||:|:|      |||      |||
orf34ng      FSGVSFRGSGWGAFFVGSTGVSLSVFSACVP-----VPVNESAARAASEGR--GLTRFFFLGA    114

orf34.pep    AGDVILLPLSSVPSGCAGSDEAAWWC SGWAASCPTT PFGSQNSVSRGLSVCCGSAXRVLS    150
|||  |||      |||      |||      |||      |||      |||      |||      |||      |||
orf34ng      AGDGSPLPLSSVPSGCAGSDEAAWWC SGWAASCPTA PFGSQNSVSRGLSVCCGSVWRVLS    174

orf34.pep    S                                                                    175

orf34ng      PFGLNVLTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD    234

```

10 The complete length ORF34ng nucleotide sequence [<SEQ ID 213>] (SEQ ID NO: 213) is:

```

1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
51  GCCGGGTCAA AAGAGGTTGT CGAGAATCTC TTTATGGGGT TTGGCCGGCG
101 TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTT
151 TCTTTGGGTG TTTCTTTGGG CTGCGCCTGT TTTTCGGGTG TTTCTTTTCG
15  201 GGGTTCGGGA TGGGGGGCGT TTGTGGGCAG TACGGGGGTT TCTTTGAGTG
251 TGTTTTCAGC TTGTGTCCG GTGCGGTTA ACGAATCGGC TGCCCGGGCC
301 GCATCCGAAG GCGCGGTTT gACCCGGTTT TTCTTGGGTG CGGCAGGGGA
351 CGGCAGTCCG CTGCCGCTTT CTTCTGTGCC GTCCGGCTGT GCGGGTTCGG
401 ATGAGGCGGC GTGGTGGTGT TCGGGTTGGG CGGCATCTTG TCCGACGGCG
20  451 CCGTTTGGCA GCCAGAATTC GGTTCGCGG GGGCTGTCCG TGTGTTGCGG
501 TTCGGTTTGG AGGGTTTTGT CGCCGTTCGG GTTGAATGTG CTGACGATGC
551 CTACTGCCAA TGCGCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
601 ATCAGGAGTT TGGGGGTCAG CCTGAAGGGT TTGTTCGGTT TTTTGGCCAT
651 TTTGATTGTG CTTTGGGGT GTCGGGCAAT GCCGTCTGAA GGCGGTTCAG
25  701 ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTTGGTAGA GGGTAATGAC
751 TTTTGTACG CCGAcgTGG TGCTGACTTT TTGGGTAATC TCGCCTGTT
801 CTTCCGGGGT GAGGATGCCC ATAACGTAGG TTACATTGCC GTAGGTAATG
851 ATTTTACGCG GCGCCTGTGT AGCGGGGCTG ATGCCAGCA GcgtgGCGCG
901 GACTTTGGAC GTGTCCAAG TGTCGCGGCG GATGTCGCCC GCAGTGCGCG
30  951 GCAGGGAGGC GACGTAATG TAGTTGTATA CGCCTTCGGC GGCCTGTTTCG
1001 GAACGTGCAA TCTGACCGAC GAACTGTTTT TCGCCTTCGG TGGCGACTTG
1051 TCCGAGCAGC AGCAGGTGGC GGTGTAGACC GACGACGGAG ATTGGGGCG
1101 TGTAGCCTTT GGTTTGGTTG TTTTGCGCA GGTAGGAACG GCGGTGGTT
1151 TCGATACGCA ACGCCATAAC GTtgtCATCG GTTtgcgcgc CGGTGGTTcg
35  1201 gCGGTCGATG ACGGATTTTG CGCCGACGGC GGCCCCGCG ACGACTGCGC
1251 TGAAGCAGCC GCCGAGGGCA AGGCTGAGGA CGGCGGCAAT CAGGGTGC GG
1301 ACGGTGTGTG GTTTGGGTTT CATCGGGGAC TTCCTTTCTT GGGCGTTTCA
1351 GACGGCATTG CTTTGCGCCA TGCCGTCTGA

```

40 This encodes a protein having amino acid sequence [<SEQ ID 214>] (SEQ ID NO: 214):

```

1  MMMPFIMLPW IAGVPAVPGQ KRLSRISLWG LAGVFFGVSG LWVFSLGVSF
51  SLGVSLGCAC FSGVSFRGSG WGAFFVGSTGV SLSVFSACVP VPVNESAARA
101 ASEGRGLTRF FLGAAGDGSP LPLSSVPSGC AGSDEAAWWC SGWAASCPTA
151 PFGSQNSVSR GLSVCCGSVW RVLSPFGLNV LTMPTANAPM AVIQMSNTAR
45  201 IRSLGVSLKG LFGFFAILIV LLGCRAMPSE GGSDGIAESA LDVVLVEGND
251 FLYADGGADF LGNLRLFFGG EDAHNVGYIA VGNDFDARLC SGADAQQRGA
301 DFGRVPVAVG DVARSAHQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL
351 SEQQVAVVA DDGDLGRVAF GLVLAQVGT GGGFDTORHN VVIGLRAGGS
401 AVDDGFCADG GPADDCAEAA AEGKAEDGNN QGADGVWFGF HRGLPFLGVS
50  451 DGIALRHAV*

```

ORF34ng (SEQ ID NO: 214) and ORF34-1 (SEQ ID NO: 210) show 90.0% identity in 459 aa overlap:

-211-

		10	20	30	40	4	50
	orf34-1.pep	MMMMFIMLPWIAGVPAVPGQNRLSRISLWGLGGVFFGVSGLVWVSLGVS-----LGCAC					
5	orf34ng	MMMMFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWVSLGVSFSLGVSLGCAC	10	20	30	40	50 60
		60	70	80	90	100	110
	orf34-1.pep	FSGVSFRGSGRGTFTVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP					
10	orf34ng	FSGVSFRGSGWGAFAVGSTGVSLSVFSACVPVPVNESAARAASEGRGLTRFFLGAAGDGSP	70	80	90	100	110 120
		120	130	140	150	160	170
	orf34-1.pep	LPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV					
15	orf34ng	LPLSSVPSGCAGSDEAAWWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLSPFGLNV	130	140	150	160	170 180
		180	190	200	210	220	230
	orf34-1.pep	LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
20	orf34ng	LTMP TANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA	190	200	210	220	230 240
		240	250	260	270	280	290
	orf34-1.pep	LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVG YVAVGNDFDARLCGGADAQQRGA					
25	orf34ng	LDVVLVEGNDFLYADGGADFLGNLRLFFGGEDAHNVG YIAVGNDFDARLCGADAQQRGA	250	260	270	280	290 300
		300	310	320	330	340	350
	orf34-1.pep	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGLFGTCNLTDDELFFAFGGDLSEQQVAVVA					
30	orf34ng	DFGRVPSVAGDVARSARQGGDGNVVVYAFGGLFGTCNLTDDELFFAFGGDLSEQQVAVVA	310	320	330	340	350 360
		360	370	380	390	400	410
	orf34-1.pep	DDGDLGRVAFGLVVLAQIGTGGGFD TQRHNVV VGLRAGGS AVDGGFRADGGASDYCADAA					
35	orf34ng	DDGDLGRVAFGLVVLAQVGTGGGFD TQRHNVV IGLRAGGS AVDDGFCADGGPADDCAEAA	370	380	390	400	410 420
		420	430	440	450		
	orf34-1.pep	AKGKAENGGNQGADGVRFGRVLPFLGVSDGIALRHAVX					
40	orf34ng	AEGKAEDGGNQGADGVWFGFHRGLPFLGVSDGIALRHAVX	430	440	450	460	

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 26

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 215>] (SEQ ID NO: 215):

```

5      1  ATGAAACCT TCTTCAAAAC CCTTTCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
     101  CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
     151  CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAAATCCAA GCCGAGCTGG
     201  AGAAAAAAGG CTACACCGTC AACTGGTTCG AGTTTACCGA CTATGTACGC
     251  CCGAATCTGG CATTGGCTGA GGGCGAGTTG

```

10 This corresponds to the amino acid sequence [<SEQ ID 216; ORF4>] (SEQ ID NO: 216; ORF4):

```

      1  MKTFFKTLA AALALILAA G.QKDSAPAA SASAAADNGA AKKEIVFGTT
     51  VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

```

Further sequence analysis revealed the complete nucleotide sequence [<SEQ ID 217>] (SEQ ID NO: 217):

```

15      1  ATGAAACCT TCTTCAAAAC CCTTTCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GGCGGTCAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
     101  CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
     151  GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
     201  GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTACGCC
20      251  CGAATCTGGC ATTGGCTGAG GGCAGATTGG ACATCAACGT CTTCCAACAC
     301  AAACCCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCAG
     351  AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
     401  AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
     451  CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
25      501  CAAACTCAAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG
     551  CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
     601  CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
     651  CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
     701  TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
30      751  TGGCTTAAAG ACGTAAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
     801  CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
     851  GCGCAGCCAA ATAA

```

This corresponds to the amino acid sequence [<SEQ ID 218; ORF4-1>] (SEQ ID NO: 218; ORF4-1):

```

35      1  MKTFFKTLA AALALILAA GGQKDSAPAA SASAAADNGA AKKEIVFGTT
      51  VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
     101  KPYLDDFKKE HNLDTTEVFQ VPTAPLGLYP GKLSLEEVEK DGSTVSAPND
     151  PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
40      201  PRSRADVDFV VVNGNYAISS GMKLTALFQ EPSFAYVNWS AVKTADKDSQ
     251  WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 (SEQ ID NO: 216) shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) (SEQ ID NO: 220) from strain A of *N. meningitidis*:

		10	20	30	40	50	59
5	orf4.pep	<u>MKTFFKTL</u> SAAALALILAACG-QKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE					
	orf4a	MKTFFKTL <u>SAAALALILAAC</u> GGQKDSAPAASASAAAADNGAAXKEIVFGTTVGDFGDMVKE					
		10	20	30	40	50	60
10	orf4.pep	60	70	80	90		
		QIQAELEKKGYTVKLVEFTDYVRPNLALAEGL					
	orf4a	XIQPELEKKGYTVKLVEXTDYVRXNLALAEGLDINVXQHXXYLDDXKKXHNLDITXVXQ					
		70	80	90	100	110	120
15	orf4a	VPTAPLGLYPGKLKSLXXVKXGSTVSAPNDPXXFXRVLVMLDELGXIKLKDIXXXXXXX					
		130	140	150	160	170	180

The complete length ORF4a nucleotide sequence [<SEQ ID 219>] (SEQ ID NO: 219) is:

	1	ATGAAAACCT	TCTTCAAAAC	CCTTTCCGCC	GCCGCACTCG	CGCTCATCCT
	51	CGCCGCCTGC	GGCGGTCAA	AAGATAGCGC	GCCCGCCGCA	TCCGCTTCTG
20	101	CCGCCGCCGA	CAACGGCGCG	GCGAANAAAG	AAATCGTCTT	CGGCACGACC
	151	GTCGGCGACT	TCGGCGATAT	GGTCAAAGAA	CANATCCAAC	CCGAGCTGGA
	201	GAAGAAAGGC	TACACCGTCA	AACTGGTCGA	GTNTACCGAC	TATGTGCGCN
	251	CGAATCTGGC	ATTGGCTGAG	GGCGAGTTGG	ACATCAACGT	CTTNCAACAC
	301	ANACNCTATC	TTGACGACTN	CAAAAAANAA	CACAATCTGG	ACATCACCNN
25	351	AGTCTTNCAA	GTGCCGACCG	CGCCTTTGGG	ACTGTACCCG	GGCAAGCTGA
	401	AATCGCTGGA	NNAAGTCAA	GANGGCAGCA	CCGTATCCGC	GCCCAACGAC
	451	CCGTNNNACT	TCGNCCGCGT	CTTGGTGATG	CTCGACGAAC	TGGGT'TNGAT
	501	CAAACTCAAA	GACNGCATCA	NNNNGNNGNN	NNNANCNANA	NNNGANANN
	551	NNNNANNNT	NNNNNNNNN	NNNNNCNCG	NNNNNNNAN	NNNNNNNNN
	601	NCGNNTNNNN	NNGCNNNNNT	NNANNNTNNN	NNCNCNNNN	NNNNNTNNNN
30	651	NANNANNAGC	GGCATGAAGC	TGACCGAAGC	CCTGTTCCAA	GAACCGAGCT
	701	TTGCCTATGT	CAACTGGTCT	GCCGTCAAAA	CCGCCGACAA	AGACAGCCAA
	751	TGGCTTAAAG	ACGTAACCGA	GGCCTATAAC	TCCGACGCGT	TCAAAGCCTA
	801	CGCGCACAAA	CGCTTCGAGG	GCTACAAATC	CCCTGCCGCA	TGGAATGAAG
35	851	GCGCAGCCAA	ATAA			

This is predicted to encode a protein having amino acid sequence [<SEQ ID 220>] (SEQ ID NO: 220):

	1	<u>MKTFFKTL</u> S	<u>AALALILAAC</u>	GGQKDSAPAA	SASAAADNGA	AXKEIVFGTT
	51	VGDFGDMVKE	XIQPELEKKG	YTVKLVEXTD	YVRXNLALAE	GELDINVXQH
40	101	XXYLDDXKKX	HNLDITXVXQ	VPTAPLGLYP	GKLKSLXXVK	XGSTVSAPND
	151	PXXFXRVLVM	LDELGXIKLK	DXIXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
	201	XXXXAXXXXX	XXXXXXXXXS	GMKLTEALFQ	EPSFAYVNWS	AVKTADKDSQ
	251	WLKDVTEAYN	SDAFKAYAHK	RFEGYKSPAA	WNEGAAK*	

45 A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence [<SEQ ID 221>]
(SEQ ID NO: 221):

```

      1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GCGCGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
5      101  CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CCGCACGACC
      151  GTCGCGGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
      201  GAAAAAAGGC TACACCGTCA AACTGGTCTGA GTTTACCGAC TATGTGCGCC
      251  CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCACAAC
      301  AAACCCCTATC TTGACGACTT CAAAAAAGAA CACAATCTGG ACATCACCGA
      351  AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
      401  AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
      451  CCGTCCAAC TCGCCGCGT CTGGGTGATG CTCGACGAAC TGGGTTGGAT
      501  CAAACTCAAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
      551  CCGAAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
      601  CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
      651  CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
      701  TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
      751  TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
      801  CGCGCACAAA CGTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
      851  GCGCAGCCAA ATAA

```

This encodes a protein having amino acid sequence [<SEQ ID 222; ORF4a-1>] (SEQ ID NO: 222; ORF4a-1):

```

      1  MKTFFKTL SA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
      51  VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
      101  KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLKSLEEVK DGSTVSAPND
      151  PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLNKI KIVELEAAQL
      201  PRSRADVDFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
      251  WLKDVTEAYN SDAFKAYAHK RFEQYKSPAA WNEGAAK*

```

ORF4a-1 (SEQ ID NO: 222) and ORF4-1 (SEQ ID NO: 218) show 99.7% identity in 287 aa overlap:

```

      10      20      30      40      50      60
orf4a-1  MKTFFKTL SAAALALILAACGGQKDSAPAA SAAADNGAAKKEIVFGTTVGDFGDMVKE
35  orf4-1  MKTFFKTL SAAALALILAACGGQKDSAPAA SAAADNGAAKKEIVFGTTVGDFGDMVKE
      10      20      30      40      50      60

      70      80      90     100     110     120
orf4a-1  QIQPELEKKGYTVKLVEFTDYVRPNLALAE GELDINVFQHKPYLDDFKKEHNLDITEVFQ
40  orf4-1  QIQAELEKKGYTVKLVEFTDYVRPNLALAE GELDINVFQHKPYLDDFKKEHNLDITEVFQ
      70      80      90     100     110     120

      130     140     150     160     170     180
orf4a-1  VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
45  orf4-1  VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
      130     140     150     160     170     180

      190     200     210     220     230     240
orf4a-1  ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
50  orf4-1  ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS

```

-215-

190 200 210 220 230 240

250 260 270 280

orf4a-1 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAX

5 orf4-1 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAX

250 260 270 280

Homology with an outer membrane protein of Pasteurella haemolítica (accession q08869) (SEQ ID NO: 1126).

ORF4 (SEQ ID NO: 216) and this outer membrane protein (SEQ ID NO: 1126) show 33% aa identity in 91aa overlap:

lip2.pasha MNFKKLLGVALVSALALTACKDEKAQAP----

15 ORF4 VXTNPDPGRTPCPSFLFETATTSGENMKTFFKLSAAAL--ALILAACGFKKTARPPHPL

110 120 130 140 150

lip2.pasha -ATTAKTENKAPLKVGVMGTGPEAQMTAVKIAKEKYGLDVELVQFTEYTQPNALHSKD

20 ORF4 LPPPTTARRKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGE

160 170 180 190 200 210

lip2.pasha LDANAFQTVPYLEQEVKDRGYKLAIIGNTLVWPIAAYSKKIKNISELKDGATVAIPNNAS

ORF4 L.....

25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 (SEQ ID NO: 216) shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) (SEQ ID NO: 224) from *N. gonorrhoeae*:

orf4nm.pep MKTFFKTL SAAALALILAACGXQKDSAPAA

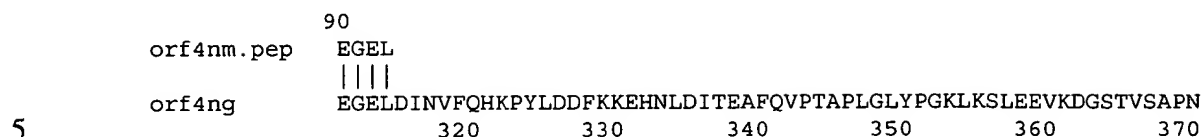
30 orf4ng RANAVXTNPDPGRTPCLSFLFETATTSGENMKTFFKTLSTASLALILAACGGQKDSAPAA

200 210 220 230 240 250

orf4nm.pep SASA-AADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA

35 orf4ng SAAAPSAADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA

260 270 280 290 300 310



The complete length ORF4ng nucleotide sequence [[SEQ ID 223](#)] ([SEQ ID NO: 223](#)) was predicted to encode a protein having amino acid sequence [[SEQ ID 224](#)] ([SEQ ID NO: 224](#)):

10
1 MKTFFKTLST ASLALILAAC GGQKDSAPAA SAAAPSadNG AAKKEIVFGT
51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ
101 HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN
151 DPSNFARALV MLNELGWIKL KGINPLTAS KADIAENLKN IKIVELEAAQ
201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
251 QWLKDVTEAY NSDAFKAYAH KRFEGYKYP A WNEGAAK*

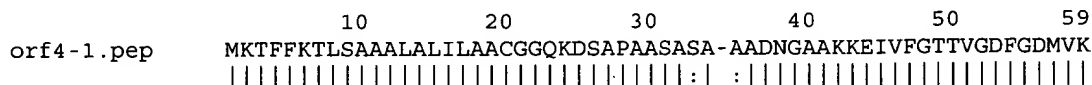
Further analysis revealed the complete length ORF4ng DNA sequence [[SEQ ID 225](#)] ([SEQ ID NO: 225](#)) to be:

20
1 atgAAAACCT TCTTCAAAAC cctttccgcc gccgcaCTCG CGCTCATCCT
51 CGCAGCCTGc ggCggtcaAA AAGACAGCGC GCCCgcagcc tctgcCGCCG
101 CCCCTTCTGC CGATAACGgc gCgGCGAAAA AAGAAAtcgt ctTCGGCACG
151 Accgtgggcg acttcggcgA TAtggTCAAA GAACAAATCC AagcCGAgct
201 gGAGAAAAAA GgctACACcg tcAAattggt cgaatttacc gactatgtGC
251 gCCCCGAATCT GGCATGGCG GAGGGCGAGT TGGACATCAA CGTCTTCCAA
301 CACAAACCTT ATCTTGACGA TTTCAAAAAA GAACACAACC TGGACATCAC
25
351 CGAAGCCTTC CAAGTGCCGA CCGCGCCTTT GGGACTGTAT CCGGGCAAAC
401 TGAATCGCT GGAAGAAGTC AAAGACGGCA GCACCGTATC CGCGCCCAac
451 gACccgTCCA ACTTCGCACG CGCCTTGGTG ATGCTGAACG AACTGGGTTG
501 GATCAAACCT AAAGACGGCA TCAATCCGCT GACCGCATCC AAAGCCGACA
551 TCGCGGAAAA CCTGAAAAAC ATCAAATCG TCGAGCTTGA AGCCGCACAA
30
601 CTGCCGCGCA GCCGCGCCGA CGTGGATTTT GCCGTCGTCA ACGGCAACTA
651 CGCCATAAGC AGCGGCATGA AGCTGACCGA AGCCCTGTTC CAAGAGCCGA
701 GCTTTGCCTA TGTCAACTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC
751 CAATGGCTTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC
801 CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG
35
851 AAGGCGCAGC CAAATAA

This encodes a protein having amino acid sequence [[SEQ ID 226; ORF4ng-1](#)] ([SEQ ID NO: 226; ORF4ng-1](#)):

40
1 MKTFFKTLSA AALALILAAC GGQKDSAPAA SAAAPSadNG AAKKEIVFGT
51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ
101 HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN
151 DPSNFARALV MLNELGWIKL KGINPLTAS KADIAENLKN IKIVELEAAQ
201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
251 QWLKDVTEAY NSDAFKAYAH KRFEGYKYP A WNEGAAK*

This shows 97.6% identity in 288 aa overlap with ORF4-1 ([SEQ ID NO: 218](#)):



5	orf4ng-1	MKTFFKTL	SAAAL	LILAACGGQKDS	SAPAASAAAPSADNGAAKKEIVFGTTVGDFGDMVK	
		10	20	30	40	50 60
10	orf4-1.pep	60	70	80	90	100 110 119
	orf4ng-1	EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFOHKPYLDDFKKEHNLDITEVF				
15	orf4-1.pep	120	130	140	150	160 170 179
	orf4ng-1	QVPTAPLGLYPGKLKLSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTAS				
20	orf4-1.pep	180	190	200	210	220 230 239
	orf4ng-1	KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW				
	orf4-1.pep	240	250	260	270	280
	orf4ng-1	SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAX				

In addition, orf4ng-1 (SEQ ID NO: 226) shows significant homology with an outer membrane protein (SEQ ID NO: 1126) from the database:

25	ID	LIP2_PASHA	STANDARD;	PRT;	276 AA.
	AC	Q08869;			
30	DT	01-NOV-1995 (REL. 32, CREATED)			
	DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)			
35	DT	01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)			
	DE	28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR. . . .			
40	SCORES	Init1: 279 Initn: 416 Opt: 494			
	Smith-Waterman score:	494; 36.0% identity in 275 aa overlap			
45	orf4ng-1.pep	10	20	30	40 50
	lip2_pasha	MKTFFKTL	SAAAL--ALILAACGGQKDS	SAPAASAAAPSADNGAAKKEIVFGTTVGDFGDM	
	orf4ng-1.pep	60	70	80	90 100 110
	lip2_pasha	VKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFOHKPYLDDFKKEHNLDITE			
	orf4ng-1.pep	120	130	140	150 160 170
	lip2_pasha	AFQVPTAPLGLYPGKLKLSLEEVKDGSTVSAPNDPSNFARVLVMLNELGWIKLKDGINPLT			

-218-

		180	190	200	210	220	230
orf4ng-1.pep		ASKADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA					
lip2_pasha		ATENDIIENPKNIKIVQADTSLLTRMLDDVELAVINNTYAGQAGLSPDKDGIIVESKDSP					
5		180	190	200	210	220	230
		240	250	260	270	280	289
orf4ng-1.pep		YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYPAAWNEGAAX					
lip2_pasha		YVNLVVSREDNKDDPRLQTFVKSFQTEEVFQEALKLFNGGVVKGW					
10		240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolítica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (SEQ ID NO: 218) (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 (SEQ ID NO: 218) is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1 (SEQ ID NO: 218).

Example 27

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 227] (SEQ ID NO: 227):

1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
101	GTTTTATGGC	TGAACCAGCA	TTATTTCCAC	GGCAACCTCC	TCTTCTACCT
151	CACCGTCGCG	ACGGCAAGCG	CACCTGGCCGG	CTGGGCGGCG	GTCGGCAAAA
201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT	GCTCATCGGC

5
10
15

```

251 GACAACGGCA GCGAATGGCT CGACAGCGGA CTCATGCGCG CCATGAACGT
301 CCTCATCGGC GyGGCCATCG CCATCGCCGC CGCCAACTG CTGCCGCTGA
351 AATCCACACT GATGTGGCGT TTCATGCTTG CCGACAACCT GGCCGACTGC
401 AGCAAAATGA TTGCCGAAAT CAGCAACGGC AGGCGCATGA CCCGCGAACG
451 CCTCGAGGAG AACATGGCGA AAATGCGCCA AATCAACGCA CGCATGGTCA
501 AAAGCCGCAG CCATCTCGCC GCCACATCGG GCGAAAGCTG CATCAGCCCC
551 GCCATGATGG AAGCCATGCA GCACGCCAC CGTAAAATCG TCAACACCAC
601 CGAGCTGCTC CTGACCACCG CCGCCAAGCT GCAATCTCCC AAACCTCAACG
651 GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....
701 ..... GC AGACACGCCG GCCGCATCCG
751 CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC
801 ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA
851 ATTTCCGCCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA
901 TGCCACGAA CGCCAACACC TGCGCCAAAG CCTGCTTGA

```

This corresponds to the amino acid sequence [<SEQ ID 228; ORF8>] (SEQ ID NO: 228; ORF8):

20

```

1 .....PRRP RHAPVSRGDL LQGGGYARH GHRAGRGFGR FMAEPALFPR
51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQT
101 HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
151 AHDPRTPRGE HGENAPNQRT HGQKPQPSRR HIGRKLHQP HDGSHAARPP
201 XNRQHHRAAP DHRRQAISQ TQRQNPAAAX PPLHTAPN...Q
251 TRPPHPHRRH HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH
301 PPQMAGCPRT PTPAPKPA*

```

25 Computer analysis of this amino acid sequence gave the following results:

Sequence motifs

ORF8 (SEQ ID NO: 228) is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

30 Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 (SEQ ID NO: 228) shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) (SEQ ID NO: 230) from *N. gonorrhoeae*:

35

```

orf8ng      1 MDRDDLRRPRHAPVPRDLLQGGGYARYGHRAGRGFGRFMAEPALFPR 50
              ||||| | | | | | | | | | | | | | | | | | | | | | | | |
orf8.pep    1 .....PRRPRHAPVSRGDLQGGGYARHGHRAGRGFGRFMAEPALFPR 44

orf8ng      51 QPLLPDHRHGKRTGRLGGGRQKRLRPYVGADDVHAHRRQRQRMARQP 100
              ||||| | | | | | | | | | | | | | | | | | | | | | | | |
orf8.pep    45 QPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQT 94

orf8ng      101 DARDERPHRRRHRHCRQTA AA E I H T D V A F H A C R Q P G R L Q Q N D C R N Q Q R Q 150
              || ||||| | | | | | | | | | | | | | | | | | | | | | | | |
orf8.pep    95 HARHERPHRRGHRHRRRQTA AA E I H T D V A F H A C R Q P G R M Q Q N D C R N Q Q R Q 144

```

40

```

orf8ng      151 AYDARTFGAEYQONAPNQRTHGQKPQPPRRHIGRKPHQPLHDGSHAARPP 200
              |:| ||   |:|:||||||||||||| ||||| ||| |||||
orf8.pep    145 AHDPRTPRGEHGENAPNQRTHGQKPQPSRRHIGRKLHQPRHDGSHAARPP 194

orf8ng      201 QNRQHHRAAPDHRRAAISQTQRQNPAAAPPLHTAPNRPATNRRPHQRQ 250
              ||||| ||||| ||||| ||||| ||||| ||||| |||||
5 orf8.pep   195 XNRQHHRAAPDHRRAAISQTQRQNPAAAPPLHTAPN.....Q 244

orf8ng      251 TRPPHPHRHRHQPRGTGSPRRTPPLPMAGFPLAQHQYASGNFRPRHPPATH 300
              ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf8.pep    245 TRPPHPHRHRHQPRGTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294

orf8ng      301 PPQMAGCPRTPTPAPKPA* 319
              ||||| ||||| ||||| ||||| |||||
10 orf8.pep   295 PPQMAGCPRTPTPAPKPA* 313

```

The complete length ORF8ng nucleotide sequence [<SEQ ID 229>] (SEQ ID NO: 229) is
 15 predicted to encode a protein having amino acid sequence [<SEQ ID 230>] (SEQ ID NO: 230):

```

1 MDRDDRLRRP RHAPVPRDL LQGGTYARY GHRAGRGFGR FMAEPALFPR
51 QPPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP
101 DARDERPHRR RHRHCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
20 orf8ng     151 AYDARTFGAE YGONAPNQR THGQKPQPPRR HIGRKPHQPL HDGSHAARPP
201 QNRQHHRAAP DHRRAAISQ TQRQNPAAAP PPLHTAPNRP ATNRRPHQRQ
251 TRPPHPHRHR HQPRGTGSPRR TPPLPMAGFP LAQHGYASGN FRPRHPPATH
301 PPQMAGCPRT PTPAPKPA*

```

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis*
 25 and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for
 raising antibodies.

Example 28

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 231>] (SEQ ID
 NO: 231):

```

30 1 ..GAAATCAGCC TGCAGTCCGA CNACAGGCCG GTTCCGTGN CGAAGCGGCG
51 GGATTCGGAA CGTTTCTGTC TGTGGACGG CGGCAACAGC CGGCTCAAGT
101 GGGCGTGGGT GGAAAACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
151 CGCGATTGT CGCCTTTGGG CGCGAGTGG GCGGAAAAGG CGGATGGAAG
201 TGTCCGCATC GTCGTTGCG CTGTGTGCGG AGAATTCAA AAGGCACAAG
35 251 TGCAGGAACA GTCGCCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
301 GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
351 CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTAGCCGC AACGCCTGCG
401 TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
451 GGACATTATC TCGGAGA.GG AACCATCATG CCCGTTTCC ACCTGATGAA
40 501 AGAATCGCTC GCCGTCCGAA CCGCAACCT CAACCGGCAC GCCGGTAAGC
551 GTTATCCTTT CCCGACCGG..

```

This corresponds to the amino acid sequence [<SEQ ID 232; ORF61>] (SEQ ID NO: 232; ORF61):

```

1  ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKWAWVENG TFATVGSAPY
51  RDLSPGLAEW AEKADGNVRI VGCAVCGEFK KAQVQEQLAR KIEWLPSSAQ
101 AXGIRNHYRH PEEHGS DRWF NALGSRRFSR NACVVVSCGT AVTVDALTTDD
151 GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..

```

Further work revealed the complete nucleotide sequence [<SEQ ID 233>] (SEQ ID NO: 233):

```

1  ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
201 TTTTCGATGCC GAAGTTTTCG GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401 GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTTTGACCG GCCGCGATAT
451 GAGTTGGGTT CGCTGTCGCC TGTTCGCGCA GTGGCGTGTC GGCGCGCCTT
501 GTCGCGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTTGGTTG
551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
601 GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
651 GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
701 GGCGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAAACGCT GTTGGTGGA
751 CTGGACGCGG TGTGTTGCA ATATGCGCGG GACGATTTG CGCCTTTTGT
801 GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
901 CAAGGCGTTT TGCATTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
951 CGCGCAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
1001 GGCGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGGTGGA AAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
1101 GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
1151 GAAATGTCCG CATCGTCGGT TGCGCTGTGT GCGGAGAATT CAAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGCGAGC GCCGCTTCAG CCGCAACGCC
1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACC GA
1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501 CGTTATCCTT TCCCACCAC AACGGCAAT GCCGTCGCCA GCGGCATGAT
1551 GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCAAAAGTTG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTACGG GTTGTGAAC ATGATTGCCG
1751 CCGAAGGCAG GGAATATGAA CATATTTAA

```

This corresponds to the amino acid sequence [<SEQ ID 234; ORF61-1>] (SEQ ID NO: 234; ORF61-1):

```

1  MTLVLKLSHWR VLAELADGLP QHVSQ LARMA DMKPQQLNGF WQQMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLREL GERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFD RPQY
151 ELGSLSPVAA VACRRALSRL GLDVQIKWPN DLVVGGRDKLG GILIETVRTG
201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLVE

```

251 L DAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVGDG
 301 QGVLHLETAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
 351 K WAWVNGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGEFKKA
 401 QVQEQLARKI EWLPSAQAL GIRNHYRHPE EHGSDFWFNA LGSRRFNRNA
 451 C VVSCGTAV TVDALTDGHL YLGGTIMPGF HLMKESLAVR TANLNRHAGK
 501 R YPFPTTTGN AVASGMMDAV CGSVMMHGR LKEKTGAGKP VDVIIITGGGA
 551 AKVAEALPPA FLAENTVRVA DNLVIYGLLN MIAAEGREYE HI*

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1 (SEQ ID NO: 234). Further computer analysis of this amino acid sequence gave the following results:

Homology with the baf protein of *B. pertussis* (accession number U12020) (SEQ ID NO: 1127).

ORF61 (SEQ ID NO: 232) and baf protein show 33% aa identity in 166aa overlap:

orf61 23 LLLDGGNSRLK WAWVE-NGTFATVGSAPYR----DLSPGLAEWA EKADGNVRIVGCAVCG 77
 +L+D GNSRLK W + + A AP DL LG A R +G V G
 baf 3 ILIDSGNSRLKVGWFDPAQAAEPAPVAFDNLDDLALGRWLATLPRRPQRALGVNVAG 62
 orf61 78 EFKKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNLGSRFRSRN 131
 + + L I WL + A G+RN YR+P++ G+DRW L +
 baf 63 LARGEIAATLRAGGCDIRWLRQAQPLAMGLRNGYRNPDLGADRWACMVGVLARQPSVHP 122
 orf61 132 ACVVVSCGTAVTVDALTDGHLGXGTIMPGFHLMKESLAVRTANL 177
 +V S GTA T+D + D + G G I+PG +M+ +LA TA+L
 baf 123 PLLVASFGTATTLDITGPDNVFPG-GLILPGPAMMRGALAYGTAHL 167

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF61 (SEQ ID NO: 232) shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) (SEQ ID NO: 236) from strain A of *N. meningitidis*:

orf61.pep 10 20 30
 EISLRSDXRPVSVKRRDSERFLLLDGGNS
 orf61a TVFEGTVKGVGDGQGVLEHLETAEGKQTVVSGEISLRSDRPVSVKRRDSERFLLLDGGNS
 290 300 310 320 330 340
 orf61.pep 40 50 60 70 80 90
 RLK WAWVNGTFATVGSAPYRDLSPGLAEWA EKADGNVRIVGCAVCGEFKKAQVQEQLAR
 orf61a RLK WAWVNGTFATVGSAPYRDLSPGLAEWA EKVDGNVRIVGCAVCGEFKKAQVQEQLAR
 350 360 370 380 390 400
 orf61.pep 100 110 120 130 140 150
 KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTD
 orf61a KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTD
 410 420 430 440 450 460
 orf61.pep 160 170 180 189
 GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT

orf61a																																																		
	GHYLG-GTIMGPFHLMKESLAVRTANLNRHAGKRYFPPTTTGNVASGMMDAVCGSVMMM																																																	
	470	480				490	500				510	520																																						
5 orf61a	HGRLKEKTGAGKPVVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG																																																	
	530	540				550	560				570	580																																						

The complete length ORF61a nucleotide sequence [<SEQ ID 235>] (SEQ ID NO: 235) is:

	1	ATGACGGTTT	TGAAGCCTTC	GCACTGGCGG	GTGTTGGCGG	AGCTTGCCGA
	51	CGGTTTGCCG	CAACACGTCT	CGCAACTGGC	GCGTATGGCG	GATATGAAGC
10	101	CGCAGCAGCT	CAACGGTTTT	TGGCAGCAGA	TGCCGGCGCA	CATACGCGGG
	151	CTGTTGCGCC	AACACGACGG	CTATTGGCGG	CTGGTGCGCC	CATTGGCGGT
	201	TTTCGATGCC	GAAGGTTTGC	GCGAGCTGGG	GGAAAGGTCG	GGTTTTTCAGA
	251	CGGCATTGAA	GCACGAGTGC	GCGTCCAGCA	ACGACGAGAT	ACTGGAATTG
	301	GCGCGGATTG	CGCCGGACAA	GGCGCACAAA	ACCATATGTG	TGACCCACCT
15	351	GCAAAGTAAG	GGCAGGGGGC	GGCAGGGGCG	GAAGTGGTCG	CACCGTTTGG
	401	GCGAGTGTCT	GATGTTTCAGT	TTTGGCTGGG	TGTTTGACCG	GCCGCAGTAT
	451	GAGTTGGGTT	CGCTGTGCGC	TGTTGCGGCA	GTGGCGTGCC	GGCGCGCCTT
	501	GTCGCGTTTG	GGTTTGAAAA	CGCAAATCAA	GTGGCCAAAC	GATTTGGTTCG
	551	TCGACGCGA	CAAATTGGGC	GGCATTCTGA	TGAAACGGT	CAGGACGGGC
20	601	GGCAAAACGG	TTGCCGTGGT	CGGTATCGGC	ATCAATTTTCG	TGCTGCCCAA
	651	GGAAGTGGAA	AACGCCGCTT	CCGTGCAATC	GCTGTTTCAG	ACGGCATCGC
	701	GGCGGGGAAA	TGCCGATGCC	GCCGTGTTGC	TGGAAACGCT	GTTGGCGGAA
	751	CTTGATGCGG	TGTTGTTGCA	ATATGCGCGG	GACGGATTTG	CGCCTTTTGT
	801	GGCGGAATAT	CAGGCTGCCA	ACCGCGACCA	CGGCAAGGCG	GTATTGCTGT
25	851	TGCGCGACGG	CGAAACCGTG	TTCGAAGGCA	CGGTTAAAGG	CGTGGACGGA
	901	CAAGGCGTTC	TGCACTTGGA	AACGGCAGAG	GGCAAAACAGA	CGGTGCTCAG
	951	CGGCGAAATC	AGCCTGCGGT	CCGACGACAG	GCCGGTTTCC	GTGCCGAAGC
	1001	GGCGGGATTG	GGAACGTTTT	CTGCTGTTGG	ACGGCGGCAA	CAGCCGGCTC
	1051	AAGTGGGCGT	GGGTGGAAAA	CGGCACGTTT	GCAACCGTCG	GTAGCGCGCC
30	1101	GTACCGCGAT	TTGTGCGCCT	TGGGCGCGGA	GTGGGCGGAA	AAGGTGGATG
	1151	GAAATGTCCG	CATCGTCGGT	TGCGCCGTGT	GCGGAGAATT	CAAAAAGGCA
	1201	CAAGTGCAGG	AACAGCTCGC	CCGAAAAATC	GAGTGGCTGC	CGTCTTCCGC
	1251	ACAGGCTTTG	GGCATACGCA	ACCACTACCG	CCACCCCGAA	GAACACGGTT
	1301	CCGACCGCTG	GTTCAACGCC	TTGGGCAGCC	GCCGCTTCAG	CCGCAACGCC
35	1351	TGCGTCGTCG	TCAGTTGCGG	CACGGCGGTA	ACGGTTGACG	CGCTCACCGA
	1401	TGACGGACAT	TATCTCGGGG	GAACCATCAT	GCCCGGTTTC	CACCTGATGA
	1451	AAGAATCGCT	CGCCGTCCGA	ACCGCCAACC	TCAACCGGCA	CGCCGGTAAG
	1501	CGTTATCCTT	TCCCGACCAC	AACGGGCAAT	GCCGTCGCCA	GCGGCATGAT
	1551	GGATGCGGTT	TGCGGCTCGG	TTATGATGAT	GCACGGGCGT	TTGAAAGAAA
40	1601	AAACCGGGGC	GGGCAAGCCT	GTCGATGTCA	TCATTACCGG	CGGCGGCGCG
	1651	GCAAAAGTTG	CCGAAGCCCT	GCCGCCTGCA	TTTTTGCGCG	AAAATACCGT
	1701	GCGCGTGGCG	GACAACCTCG	TCATTACCGG	GCTGCTGAAC	CTGATTGCCG
	1751	CCGAAGGCGG	GGAATCGGAA	CATACTTAA		

45 This encodes a protein having amino acid sequence [<SEQ ID 236>] (SEQ ID NO: 236):

	1	MTVLKPSHWR	VLAELADGLP	QHVSQALARMA	DMKPQQLNGF	WQMPAHIRG
	51	LLRQHDGYWR	LVRPLAVFDA	EGLRELGERS	GFQTALKHEC	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HLRGECLMFS	FGWVFDPRQY
50	151	ELGSLSPVAA	VACRRALSRL	GLKTQIKWPN	DLVVGRDKLG	GILIETVRTG
	201	GKTAVAVGIG	INFVLPKEVE	NAASVQSLFQ	TASRRGNADA	AVLLETLLAE
	251	LDAVLLQYAR	DGFAPFVAEY	QAANRDHGKA	VLLLRDGETV	FEPTVKGVVDG
	301	QGVHLHLETAE	GKQTVVSGEI	SLRSDDRPVS	VPKRRDSERF	LLLDGGNSRL
	351	KWAVVENGTF	ATVGSAPYRD	LSPLGAEWAE	KVDGNVRIVG	CAVCGEFKKA
	401	QVQEQLARKI	EWLPSAQAL	GIRNHYRHPE	EHGSDRWFNA	LGSRRFNRNA
55	451	CVVVSCTGAV	TVDALTDGHH	YLGGTIMPGF	HLMKESLAVR	TANLNRHAGK
	501	RYPFPTTTGN	AVASGMMDAV	CGSVMMMHGR	LKEKTGAGKP	VDVIITGGGA

551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

ORF61a (SEQ ID NO: 236) and ORF61-1 (SEQ ID NO: 234) show 98.5% identity in 591 aa overlap:

```

5      10      20      30      40      50      60
orf61a.pep  MTVLKPSHWRVLAELADGLPQHVSQ LARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR
orf61-1      MTVLKLSHWRVLAELADGLPQHVSQ LARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR
           10      20      30      40      50      60

10     70     80     90    100    110    120
orf61a.pep  LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
orf61-1      LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
           70     80     90    100    110    120

15     130    140    150    160    170    180
orf61a.pep  GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAACRRALSRLGLKTQIKWPN
orf61-1      GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAACRRALSRLGLDVQIKWPN
           130    140    150    160    170    180

20     190    200    210    220    230    240
orf61a.pep  DLVVGRDKLGGILIE TVRTGGKTVA VVGIGINFVLPKEVENAASVQSLFQTASRRGNADA
orf61-1      DLVVGRDKLGGILIE TVRTGGKTVA VVGIGINFVLPKEVENAASVQSLFQTASRRGNADA
           190    200    210    220    230    240

25     250    260    270    280    290    300
orf61a.pep  AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG
orf61-1      AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG
           250    260    270    280    290    300

30     310    320    330    340    350    360
orf61a.pep  QGV LHLETAEGKQTVVSGEISLRSDRPVSVPKRRDSERFLLDGGNSRLKWAWVENGTF
orf61-1      QGV LHLETAEGKQTVVSGEISLRSDRPVSVPKRRDSERFLLDGGNSRLKWAWVENGTF
           310    320    330    340    350    360

35     370    380    390    400    410    420
orf61a.pep  ATVGSAPYRDLSP LGAEWA EKVDGNVRIVGCAVCGEFKKAQVQEQ LARKIEWLPSSAQAL
orf61-1      ATVGSAPYRDLSP LGAEWA EKADGNVRIVGCAVCGEFKKAQVQEQ LARKIEWLPSSAQAL
           370    380    390    400    410    420

40     430    440    450    460    470    480
orf61a.pep  GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHYLG GTIMPGF
orf61-1      GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHYLG GTIMPGF
           430    440    450    460    470    480

45     490    500    510    520    530    540
orf61a.pep  HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP
orf61-1      HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP

```


-225-

		490	500	510	520	530	540
		550	560	570	580	590	
	orf61a.pep	VDVIITGGGA	AKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHTX				
5	orf61-1	VDVIITGGGA	AKVAEALPPAFLAENTVRVADNLVIYGLLNLMIAAEGREYEHIX				
		550	560	570	580	590	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 (SEQ ID NO: 232) shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) (SEQ ID NO: 238) from *N. gonorrhoeae*:

10	orf61.pep	EISLRSDXRPVS	VXKRRD	SERFLLLDGGNS	30		
	orf61ng	TVCEGTVKGV	DGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPD	SERFLLLEGGNS	211		
15	orf61.pep	RLKWAVVENGT	FATVGSAPYRDLSP	LGAEWA	EKADGNVRI	VGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKWAVVENGT	FATVGSAPYRDLSP	LGAEWA	EKADGNVRI	VGCAVCGESKKAQVQEQLAR	271
	orf61.pep	KIEWLPSSAQ	AXGIRNHYRHPEEHGSDRWF	NALGSRRF	SRNACVVVSCGTAVTVDAL	TDD	150
	orf61ng	KIEWLPSSAQ	ALGIRNHYRHPEEHGSDRWF	NALGSRRF	SRNACVVVSCGTAVTVDAL	TDD	331
20	orf61.pep	GHYLGXGT	IMP	GFHLMKESLAVRTANLNRHAGKRYPFPT			189
	orf61ng	GHYLG-GT	IMP	GFHLMKESLAVRTANLNR	PAGKRYPFPTTTGNAVASGMM	DAVCGSIMMM	390

An ORF61ng nucleotide sequence [<SEQ ID 237>] (SEQ ID NO: 237) was predicted to encode a protein having amino acid sequence [<SEQ ID 238>] (SEQ ID NO: 238):

25	1	MFSFGWAFDR	PQYELGSLSP	VAALACRRAL	GCLGLETQIK	WPNDLVVGRD
	51	KLGGILIETV	RAGGKTVA	VV GIGINFLPK	EVENAASVQS	LFQTASRRGN
	101	ADAAVLLETL	LAELGAVLEQ	YAEFGAPFL	NEYETANRDH	GKAVLLLRDG
	151	ETVCEGTVKG	VDGRGVLHLE	TAEGEQTVVS	GEISLRPDNR	SVSVKRPDS
30	201	ERFLLLEGGN	SRLKWAVVEN	GTFATVGSAP	YRDLSP	LGAE WAEKADGNVR
	251	IVGCAVCGES	KKAQVKEQLA	RKIEWLPSSA	QALGIRNHYR	HPEEHGSDRW
	301	FNALGSRRFS	RNACVVVSCG	TAVTVDALTD	DGHYLG	GTIM PGFHLMKESL
	351	AVRTANLNR	PAGKRYPFPTT	TGNAVASGMM	DAVCGSIMMM	HGRLKEKNGA
	401	GKPV	VDVIITG	GGA	AKVAEAL	PPAFLAENTV RVADNLVIHG LLNLIAAEGG
35	451	ESEHA*				

Further analysis revealed the complete gonococcal DNA sequence [<SEQ ID 239>] (SEQ ID NO: 239) to be:

40	1	ATGACG	GTTT	TGAAGC	CTTC	GCATTG	GCGG	GTGTTG	GCGG	AGCTTG	CCGA
	51	CGGTTT	GCCG	CAACAC	GTAT	CGCAAT	TGGC	GCGTGAG	GCG	GACATGA	AGC
	101	CGCAGC	AGCT	CAACG	GTTTT	TGGCAG	CAGA	TGCCG	GCGCA	TATACG	CGGG
	151	CTGTTG	CGCC	AACACG	ACGG	CTATTG	GCGG	CTGGTG	CGCC	CCTTGG	CGGT
	201	TTTCGAT	GCC	GAAGGT	TTGC	GCGATC	TGGG	GGAAGG	TCG	GGTTTT	CAGA
	251	CGGCAT	TGAA	GCACG	AGTGC	GCGTCC	AGCA	ACGACG	AGAT	ACTGGA	ATTG

```

5   301 GCGCGGATTG CGCCGGACAA GCGGCACAAA ACCATATGCG TGACCCACCT
    351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
    401 GCGAGTGCCT GATGTTTCA GTTCGGCTGGG CGTTTGACCG GCCGCAGTAT
    451 GAGTTGGGTT CGCTGTGCGC TGTTCGCGCA CTTGCGTGCC GGCGCGCTTT
    501 GGGGTGTTTG GGTTCGAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
    551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGGCGGGC
    601 GGTAAACCGG TTGCCGTGGT CGGTATCGGC ATCAATTTTC TGCTGCCCAA
    651 GGAAGTGGAA AACGCCGCTT CCGTGCAGTC GCTGTTTCAG ACGGCATCGC
    701 GCGGGGGCAA TGCCGATGCC GCCGTATTGC TGGAAACATT GCTTGCGGAA
10  751 CTGGGCGCGG TGTTCGAAAC ATATGCGGAA GAAGGGTTCG CGCCATTTTTT
    801 AAATGAGTAT GAAACGGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
    851 TGC GCGACGG CGAAACCGTG TCGGAAGGCA CGGTTAAAGG CGTGGACGGA
    901 CGAGCGGTTT TGCACCTTGA AACGGCAgaa ggcgaACAGa cggtcgctcag
    951 cggcgaaaTC AGcctGCggc ccgacaacaG GTCGGtttcc gtgccgaagc
15 1001 ggccggatTC GgaacgtTTT tTGctgttgg aaggcgggaa cagccgGCTC
    1051 AAGTGGGCGT GggtggAAAA cggcacgttc gcaaccgtgg gcagcgcgCc
    1101 gtaCCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
    1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATC CAAAAAGGCA
    1201 CAAGTGAAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
20 1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
    1301 CCGACCGTTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
    1351 TGCCTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
    1401 TGACGGACAT TATCTCGGCG GAACCATCAT GCCCGGCTTC CACCTGATGA
    1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
25 1501 CGTTACCCTT TCCCGACCAC AACGGGCAAC GCCGTCGCAA GCGGCATGAT
    1551 GGACGCGGTT TCGGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
    1601 AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
    1651 GCGAAAGTCG CCGAAGCCCT GCCGCCTGCA TTTTGGCGG AAAATACCGT
    1701 GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
30 1751 CCGAAGGCGG GGAATCGGAA CACGCTTAA

```

This corresponds to the amino acid sequence [SEQ ID 240; ORF61ng-1] (SEQ ID NO: 240; ORF61ng-1):

```

35 1   MTVLKPSHWR VLAELADGLP QHVSQ LAREA DMKPQQLNGF WQMPAHIRG
    51 LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL
    101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWAFDRPQY
    151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILIETVRAG
    201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
40 251 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVVDG
    301 RGVLHLETAE GEQTVVSGEI SLRPDNRVS VPKRPDSERF LLEGGNSRL
    351 KWAUVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGESKKA
    401 QVKEQLARKI EWLPSAQAAL GIRNHYRHPE EHGS DRWFNA LGSRRF SRNA
    451 CUVVSCGTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANLNR PAGK
    501 RYFPPTTGN AVASGMMDAV CGSIMMHGR LKEKNGAGKP VDVIITGGGA
45 551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA*

```

ORF61ng-1 (SEQ ID NO: 240) and ORF61-1 (SEQ ID NO: 234) show 93.9% identity in 591 aa overlap:

```

50 orf61ng-1.pep  MTVLKPSHWRVLAELADGLPQHVSQ LAREADMKPQQLNGFWQMPAHIRG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf61-1       MTVLKLSHWRVLAELADGLPQHVSQ LARMADMKPQQLNGFWQMPAHIRG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

    orf61ng-1.pep  LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK	120
	orf61ng-1.pep	GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN	180
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN	180
5	orf61ng-1.pep	DLVVGRDKLGGILIIETVRAGGKTAVVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61-1	DLVVGRDKLGGILIIETVRTGGKTAVVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG	300
10	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEQTVKGVDG	300
	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGNSRLKWAVVENGT	360
	orf61-1	QGVHLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGNSRLKWAVVENGT	360
15	orf61ng-1.pep	ATVGSAPYRDLSPPLGAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL	420
	orf61-1	ATVGSAPYRDLSPPLGAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	420
	orf61ng-1.pep	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGTTIMPGF	480
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGTTIMPGF	480
20	orf61ng-1.pep	HLMKESLAVRTANLNRPAKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP	540
	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP	540
	orf61ng-1.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHAX	593
25	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNLMIAAEGREYEHIX	593

Based on this analysis, including the homology with the baf protein (SEQ ID NO: 1127) of *B.pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 29

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 241] (SEQ ID NO: 241):

	1	ATGTTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
35	101	GCCTGCTAAT	TGCCGCGCTG	CCTGCACTGC	CCGCCTGCCG	CCGTCATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTTCGT
	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTGCGGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT

5 351 ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
 401 CGGaAGAGGG CGGCGaAGTC GGCTGGTTCG GCTGCCTGCT GGTGTGTGTTG
 451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
 551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGC..

This corresponds to the amino acid sequence [<SEQ ID 242; ORF62>] (SEQ ID NO: 242; ORF62):

10 1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 201 WSVGMVLSLL YLGLGC..

15 Further work revealed the complete nucleotide sequence [<SEQ ID 243>] (SEQ ID NO: 243):

 1 ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
 101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
 151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCTG
 201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
 251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
 301 TTTGTCGGAC ACTTTTCTCT CAACGACAAA GCGCGTGCCT ACCACTGGAT
 351 ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
 401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTGTGTTG
 451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
 551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGC GG
 651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCTGCCA
 701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTCTGCTG
 751 GCGGTTTTGA TTTTGGGCGA ACACCTGTCT CCCGTGTCGG CCTTGGGCGT
 801 GTTTGTCGTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAAA
 851 AATAA

35 This corresponds to the amino acid sequence [<SEQ ID 244; ORF62-1>] (SEQ ID NO: 244; ORF62-1):

 1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 201 WSVGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANVSGLLI SLEPVVGVL
 251 AVLILGEHLS PVSALGVFVV IAATLVAGRL SHQK*

Computer analysis of this amino acid sequence gave the following results:

45 Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147) (SEQ ID NO: 1128)

ORF62 (SEQ ID NO: 242) and HI0976 (SEQ ID NO: 1128) show 50% aa identity in 114aa overlap:

```

Orf62   1  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXCRRHVGKIPREEWKP 60
5        M YQILAL+IWSSS I K Y +DP L+V VR R KI + K
HI0976  1  MLYQILALLIWSSSLIVGKLTYSMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60

Orf62   61  LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAY 114
        L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFGHHFF K +
HI0976  61  LWWLAFFNYTAVFLQFIGLKYTSAASAVTMIGLEPLLVMFVGHHFFKTKQNGF 114

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF62 (SEQ ID NO: 242) shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) (SEQ ID NO: 246) from strain A of *N. meningitidis*:

```

15 orf62.pep  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGKIPREEWKP
orf62a      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGKIPREEWKP
           10      20      30      40      50      60

orf62.pep  LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAYHWICGA
20 orf62a    LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAYHWICGA
           70      80      90     100     110     120

orf62.pep  AAFAGVALLMAGGAEEGGEVGFCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA
25 orf62a    AAFAGVALLMAGGAEEGGEVGFCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA
           130     140     150     160     170     180

orf62.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
30 orf62a    AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI
           190     200     210     220     230     240

orf62a      SLEPVVGVLAVLILGEHLSPVSVLGVFVIAATLVAGRLSHQKX
           250     260     270     280

```

35 The complete length ORF62a nucleotide sequence [<SEQ ID 245>] (SEQ ID NO: 245) is:

```

1  ATGTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101 GCCTGCTGAT TGCTGCGCTG CCTGCACTGC CCGCTGCCG CCGTCATGTC
151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCTG
40 201 CAACTATGTG CTGACCTTGC TACTTCAGTT TGTCGGGTTG AAATACACTT
251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCACT GCTGATGGTG
301 TTTGTGCGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351 ATGCGGCGCG GCGGCATTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG

```

-230-

5
10

```

401  CGGAAGAGGG  CGGCGAAGTC  GGCTGGTTCG  GCTGCCTGCT  GGTGTTGTTG
451  GCGGGCGCGG  GCTTTTGTGC  CGCTATGCGT  CCGACGCAAA  GGCTGATTGC
501  ACGCATCGGC  GCACCGGCAT  TCACATCTGT  TTCCATTGCC  GCCGCATCGT
551  TGATGTGCCT  GCCGTTTTTC  CTTGCTTTGG  CGCAAAGTTA  TACCGTGGAC
601  TGGAGCGTCG  GAATGGTATT  GTCGCTGCTG  TATTTGGGCG  TGGGGTGCAG
651  CTGGTACGCC  TATTGGCTGT  GGAACAAGGG  GATGAGCCGT  GTTCCTGCCA
701  ACGTTTCGGG  ACTGTTGATT  TCGCTCGAAC  CCGTCGTCGG  CGTGCTGCTG
751  GCGGTTTTGA  TTTTGGGCGA  ACACCTGTGC  CCCGTGTCCG  TCTTGGGCGT
801  GTTTGTCTGC  ATCGCCGCCA  CCTTGGTTGC  CGGCCGGCTG  TCGCATCAAA
851  AATAA

```

This encodes a protein having amino acid sequence [SEQ ID 246] (SEQ ID NO: 246):

15

```

1   MFYQILALII  WSSSFIAAKY  VYGGIDPALM  VGVRLIIAAL  PALPACRRHV
51  GKIPREEWKP  LLIVSFVNYV  LTLLLQFVGL  KYTSAASASV  IVGLEPLLMV
101 FVGHFFFNDK  ARAYHWICGA  AAFAGVALLM  AGGAEEGGEV  GWFGLLVLL
151 AGAGFCAAMR  PTQRLIARIG  APAFTSVSIA  AASLMCLPFS  LALAQSYTVD
201 WSVGMLSL  YLGVGCSWYA  YWLWNKGMSR  VPANVSGLLI  SLEPVVGVL
251 AVLILGEHLS  PVSVLGVFVV  IAATLVAGRL  SHQK*

```

20 ORF62a (SEQ ID NO: 246) and ORF62-1 (SEQ ID NO: 244) show 98.9% identity in 284 aa overlap:

25

```

orf62a.pep  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV 60
|
orf62-1      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV 60
|
orf62a.pep  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFN 120
|
orf62-1      LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFN 120
|
orf62a.pep  AAFAGVALLMAGGAEEGGEVGVGFCCLLVLLAGAGFCAAMRPTQRLIAR 180
|
orf62-1      AAFAGVALLMAGGAEEGGEVGVGFCCLLVLLAGAGFCAAMRPTQRLIAR 180
|
orf62a.pep  AASLMCLPFSLALAQSYTVDWSVGMLSLLYLGVGCSWYAYWLWNKGMSR 240
|
orf62-1      AASLMCLPFSLALAQSYTVDWSVGMLSLLYLGLGCGWYAYWLWNKGMSR 240
|
orf62a.pep  SLEPVVGVLAVLILGEHLS 285
|
orf62-1      SLEPVVGVLAVLILGEHLS 285

```

35

Homology with a predicted ORF from *N. gonorrhoeae*

ORF62 (SEQ ID NO: 242) shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) (SEQ ID NO: 248) from *N. gonorrhoeae*:

40

```

orf62.pep  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV 60
|
orf62ng     MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV 60
|

```

-231-

orf62.pep LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHFFNDKARAYHWICGA 120
|||
orf62ng LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHFFNDKARAYHWICGA 120
|||
orf62.pep AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA 180
|||
orf62ng AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA 180
|||
orf62.pep AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC 216
|||
orf62ng AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI 240
|||

The complete length ORF62ng nucleotide sequence [<SEQ ID 247>] (SEQ ID NO: 247) is:

1 ATGTTTACC AAATCCTGCT CCTGATTATC TGGGGCAGCT CGTTTATTGC
51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101 GCCTGCTGAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCTG
201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
251 CCGCCGCCAG CGCATCGGTC ATTGTCCGAC TCGAGCCGCT GCTGATGGTG
301 TTTGTCCGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351 ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
501 CCGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
601 TGGAGCGTCG GGATGGTATT GTCGCTGTTG TATTTGGGTT TGGGGTCCG
651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCTGCCA
701 ACGCGTCGGG ACTGTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGTTG
751 GCGGTTTTGA TTTTGGGCGA ACATTATCG CCCGTGTCCG CCTTGGGCGT
801 GTTTGTGCTC ATCGCCGCCA CTTTCGCCGC CGGCCGGCTG TCGCGCAGGG
851 ACGCGCAAAA CGCAATGCC GTCTGA

This encodes a protein having amino acid sequence [<SEQ ID 248>] (SEQ ID NO: 248):

1 MFYQILALII WGSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
101 FVGHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGLLVLL
151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201 WSVGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVL
251 AVLILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*

ORF62ng (SEQ ID NO: 248) and ORF62-1 (SEQ ID NO: 244) show 97.9% identity in 283 aa overlap:

10 20 30 40 50 60
orf62ng.pep MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV
|||
orf62-1 MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV
|||
10 20 30 40 50 60
70 80 90 100 110 120
orf62ng.pep LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHFFNDKARAYHWICGA
|||
orf62-1 LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFGHFFNDKARAYHWICGA
|||
70 80 90 100 110 120

-232-

5 orf62ng.pep 130 140 150 160 170 180
AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA
orf62-1 130 140 150 160 170 180
AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA

10 orf62ng.pep 190 200 210 220 230 240
AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI
orf62-1 190 200 210 220 230 240
AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI

15 orf62ng.pep 250 260 270 280 290
SLEPVVGVLAVLILGEHLSPVSA LGVFVVIATFAAGRLSRRDAQNGNAVX
orf62-1 250 260 270 280
SLEPVVGVLAVLILGEHLSPVSA LGVFVVIATLVAGRLSHQKX

Furthermore, ORF62ng (SEQ ID NO: 248) shows significant homology to a hypothetical *H.influenzae* protein (SEQ ID NO: 1128):

20 sp|Q57147|Y976_HAEIN HYPOTHETICAL PROTEIN HI0976)gi|1074589|pir||B64163
hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
)gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
Score = 106 bits (262), Expect = 2e-22
Identities = 56/114 (49%), Positives = 68/114 (59%)

25 Query: 1 MFYQILALI IWGSSFIAAKYVYGGIDPALMVGVRXXXXXXXXXXCRRHVVGKIPREEWKP 60
M YQILAL+IW SS I K Y +DP L+V VR R KI + K
Sbjct: 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVQVRLIAMIIVMPLFLRRWKKIDKPMRKQ 60

Query: 61 LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLL MVFVGHHFFNDKARAY 114
L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHHFF K +
Sbjct: 61 LWWLAFFNYTAVFLLQFIGLKYTSASSAVTMIGLEPLL VVVFVGHHFFKTKQNGF 114

30 Based on this analysis, including the homology with the transmembrane protein (SEQ ID NO: 1128) of *H.influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 30

35 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 249>] (SEQ ID NO: 249):

40 1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkgTA
51 sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
101 GGTGGATTGT TGC GTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
201 CGGTTTCGCtA srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGTTT

5
10
15
20

```

251  GCCGkACTGC  CCGGCGTGTT  TCTGTTCCGGC  TTTCCCGCAC  AGTTCATCAA
301  CGGCACGATT  AATTCGTGGT  TCGGCAACGA  TACCCACGAG  GCGCTTGAAC
351  GCAGCCTCAA  TTTGAGCAAG  TCCGCATTGA  ATTTGGCGGC  AGACAACGCC
401  CTCGGCAACG  CCGTCCCCGT  GCAGATAGAC  CTCATCGGCG  CGGCTTCCCT
451  GCCCAGGGAT  ATGGGCAGGG  TGCTGGAACA  TTACGCCGGC  AGCGGTTTTG
501  CCCAGCTTGC  CCTGTACAAy  ksCGCAAGCG  GCAAAATCGA  AAAAAGCATC
551  AACC CGCACA  AGCTCGATCA  GCCGTTTCCA  GGTAAGGCGC  GTTGGGAaAa
601  AATCCaACGG  GCGGGTTCGG  TCAGGGATTT  GGAAAGCATA  GGCGGCGTAT
651  TGTaCGCGCA  GGGCTGGCTG  TCGGCGGGTA  CGCACwACGG  GCGGATTAC
701  GCCTTGTTTT  TCCGTCAGCC  GGTTCCTCAA  GGCGTGCGAG  AGGATGCCGT
751  yTTAAATCGAA  AAGGCAAGGG  CGAAATATGC  TGAGTTGAGT  TACAGCAAAA
801  AAGGTTTGCA  GACCTTTTTT  CTGGCAACCC  TGCTGATTGC  CTCGCTGCTG
851  TCGATTTTTT  TTGCACTGGT  CATGGCACTG  TATTTTCGCC  GCCGTTTCGT
901  CGAACCCGTC  CTATCGCTTG  CCGAGGGGGC  GAAGGCGGTG  GCGCAAGGCG
951  ATTTAGCCA  GACGCGCCCC  GTGTTGCGCA  ACGACGAGTT  CGGACGCTTG
1001 ACCArGTTGT  TCAACCAT  GACCGAGCAG  CTTTCCATCG  CCAAAGATGC
1051 AGACGAGCGC  AACCGCCGGC  GCGAGGAAGC  CGCCAGGCAT  TATCTTGAAT
1101 GCGTGTTGGA  GGGGCTGACC  ACGGGCGTGG  TGGTGTTTGA  CGAACAAGGC
1151 TGTCTGAAAA  CCTTCAACAA  AGCGGCGGGT  ACC. .

```

This corresponds to the amino acid sequence [[SEQ ID 250](#); [ORF64](#)] ([SEQ ID NO: 250](#); [ORF64](#)):

25
30

```

1  MRRFLPIAAI  CAXXLXXGLT  AATGSTSSLA  DYFWWIVAFS  AMLLLVLSAV
51  LARYVILLK  DRRDGVFGSX  XAKXPXXMF  TLVAXLPGVF  LFGFPAQFIN
101  GTINSWFND  THEALERSLN  LSKSALNLAA  DNALGNAPV  QIDLIGAASL
151  PGDMGRVLEH  YAGSGFAQLA  LYNXASGKIE  KSNIPHKLDQ  PFPKGARWEK
201  IQRAGSVRDL  ESIGGVLYAQ  GWLSAGTHXG  RDYALFFRQP  VPKGVAEDAV
251  LIEKARAKYA  ELSYSKGLQ  TFFLATLLIA  SLLSIFLALV  MALYFARRFV
301  EPVLSLAEGA  KAVAQGDfsQ  TRPVLNRDEF  GRLTXLFNHM  TEQLSIAKDA
351  DERNRRREEA  ARHYLECVLE  GLTTGVVVF  EQGCLKTFNK  AAGT. .

```

Further work revealed the complete nucleotide sequence [[SEQ ID 251](#)] ([SEQ ID NO: 251](#)):

35
40
45
50
55

```

1  ATGCGCCGTT  TTCTACCGAT  CGCAGCCATA  TGCGCCGTCG  TCCTGTTGTA
51  CGGACTGACG  GCGCAACCG  GCAGCACCAG  TTCGCTGGCG  GATTATTTCT
101  GGTGGATTGT  TCGGTTGAGC  GCAATGCTGC  TGCTGGTGTT  GTCCGCGGTT
151  TTGGCAGGTT  ATGTCATATT  GCTGTTGAAA  GACAGGCGCG  ACGGCGTATT
201  CGGTTGCGAG  ATTGCCAAAC  GCCTTCTGCG  GATGTTTACG  CTGGTTGCCG
251  TACTGCCCGG  CGTGTTTCTG  TTCGCGGTTT  CCGCACAGTT  CATCAACGGC
301  ACGATTAATT  CGTGGTTCGG  CAACGATACC  CACGAGGCGC  TTGAACGCAG
351  CCTCAATTG  AGCAAGTCCG  CATTGAATTT  GGCGGCAGAC  AACGCCCTCG
401  GCAACGCCGT  CCCCCTGAG  ATAGACCTCA  TCGGCGCGGC  TTCCCTGCCC
451  GGGGATATGG  GCAGGGTGCT  GGAACATTAC  GCCGGCAGCG  GTTTTGCCCA
501  GCTTGCCCTG  TACAATGCCG  CAAGCGGCAA  AATCGAAAAA  AGCATCAACC
551  CGCACAAGCT  CGATCAGCCG  TTTCCAGGTA  AGGCGCGTTG  GGAAAAAATC
601  CAACGGGCGG  GTTCGGTCAG  GGATTTGGAA  AGCATAGGCG  GCGTATTGTA
651  CGCGCAGGGC  TGGCTGTGCG  CGGGTACGCA  CAACGGGCGC  GATTACGCCT
701  TGTTTTTCCG  TCAGCGGTT  CCCAAAGGCG  TGGCAGAGGA  TGCCGTCTTA
751  ATCGAAAAAG  CAAGGGCGAA  ATATGCTGAG  TTGAGTTACA  GCAAAAAAGG
801  TTTGCAGACC  TTTTCTCTGG  CAACCCTGCT  GATTGCCTCG  CTGCTGTCTG
851  TTTTCTTGC  ACTGGTCATG  GCACTGTATT  TCGCCCGCCG  TTTGTCGAA
901  CCCGTCTAT  CGCTTGCCGA  GGGGCGGAAG  GCGGTGGCGC  AAGGCGATTT
951  CAGCCAGACG  CGCCCCGTGT  TGCGCAACGA  CGAGTTTCGA  CGCTTGACCA
1001 AGTTGTTCAA  CCACATGACC  GAGCAGCTTT  CCATCGCCAA  AGAAGCAGAC
1051 GAGCGCAACC  GCCGCGCGCA  GGAAGCCGCC  AGGCATTATC  TTGAATGCGT
1101 GTTGAGGGG  CTGACCACGG  GCGTGGTGGT  GTTTGACGAA  CAAGGCTGTC
1151 TGA AACCTT  CAACAAAGCG  GCGGAACAGA  TTTTGGGGAT  GCCGCTTACC

```

5
10
15
20

```

1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301 ACAAACCGGT CCAATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CAACCGTCTT GCCCGAAGAC AACGGCAACG GCGTGGTAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
1601 AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCGG CAATTATGCG
1651 CGTTCCCTT CGCTCAAAAT GGAAATCAG GATTGAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
1751 TTGCCGCGA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TGTGCCCGAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTCGGATTG
1901 TCCTGACGGT TTGCGACAAC GGCAAAGGTT TCGGCAGGGA AATGCTGCAC
1951 AACGCCTTCG AGCCGTATGT AACGACAAA CCGCGGGGAA CGGGATTGGG
2001 TCTGCCGTGT GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
2051 TGAGCAATCA GGATGCGGGT GGCGCGTGTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAAAAA CTTATGCGTA G

```

This corresponds to the amino acid sequence [<SEQ ID 252; ORF64-1>] (SEQ ID NO: 252; ORF64-1):

25
30
35

```

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51 LARYVILLK DRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAVVPQ IDLIGAASLP
151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QRAGSVRDLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAKYAE LSYSKKGLOT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301 PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401 PLWGSSRHGW HGVSAQQSLL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL
451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA
551 RSPSLKLENQ DLNALIGDVL ALYEAGPCRF AEELAGEPLT VAADTTAMRQ
601 VLHNI FKNAA EAAEEADVPE VRVKSETGQD GRIVLTVCDN GKGFGREMLH
651 NAFEPYVTDK PAGTGLGLPV VKKIIIEHGG RISLSNQDAG GACVRIILPK
701 TVKTYA*

```

Computer analysis of this amino acid sequence gave the following results:

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 (SEQ ID NO: 250) shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) (SEQ ID NO: 254) from strain A of *N. meningitidis*:

45

```

          10      20      30      40      50      60
orf64 .pep MRRFLPIAAICAXLXXGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
          |||||  |||||  |||||  |||||  |||||  |||||
orf64a     MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
          10      20      30      40      50      60

          70      80      90     100     110     120
orf64 .pep DRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGPAQFINGTINSWFGNDTHEALERSLN

```

	orf64a	DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFVGSAQFINGTINSWFGNDTHEALERSLN 70 80 90 100 110
5	orf64.pep	LSKSALNLAADNALGNAVVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE 130 140 150 160 170 180
	orf64a	LSKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIE 120 130 140 150 160 170
10	orf64.pep	KSINPHKLDQPFPKGARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP 190 200 210 220 230 240
	orf64a	KSINPHKLDQPFPKGARWEKIQQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFFRQP 180 190 200 210 220 230
15	orf64.pep	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV 250 260 270 280 290 300
	orf64a	VPKGVAEDAVLIEKARAXXXLSYSKKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV 240 250 260 270 280 290
20	orf64.pep	EPVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTXLFNHMTQEQLSIAKDADERNNRRREEA 310 320 330 340 350 360
	orf64a	EPVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMTQEQLSIAKEADERNNRRREEA 300 310 320 330 340 350
25	orf64.pep	ARHYLECYLEGLTTGVVVVFDEQGCLKT FNKAAGT 370 380 390
	orf64a	ARHYLECYLEGLTTGVVVVFDEQGCLKT FNKAAEQILGMPLTPLWGSSRHGWHGVSAAQSSL 360 370 380 390 400 410
30	orf64a	LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMIDDITVLIHAQ 420 430 440 450 460 470

The complete length ORF64a nucleotide sequence [<SEQ ID 253>] (SEQ ID NO: 253) is:

	1	ATGCGCCGTT	TTCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGTTGTA
35	51	CGGACTGACG	GCGGCAACCG	GCAGACCAG	TTCTGCTGGC	GATTATTTCT
	101	GGTGGATTGT	TGCGTTCAGC	GCAATGCTGC	TGCTGGTGT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCG	ACGGCGTATT
	201	CAGTTTCGAG	ATTGCCAAAC	GCCTTTCCGG	GATGTTTACG	CTGGTTGCCG
	251	TACTGCCCGG	CGTGTTCGT	TTCTGGCGTTT	CCGCACAGTT	TATCAACGGC
40	301	ACGATTAATT	CGTGGTTCGG	CAACGATACC	CACGAGGCGC	TTGAACGCAG
	351	CCTCAATTTG	AGCAAGTCCG	CATTGAATCT	GGCGGCAGAC	AACGCCCTTG
	401	GCAACGCCAT	CCCCGTGCAG	ATAGACNTCA	TCGGCGCGGC	TTCCCTGCCC
	451	NGGGATATGG	GCAGGGTGCT	GGAACATTAC	GCCGCGAGCG	GTTTTGCCCA
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGCAA	AATCGAAAAA	AGCATCAAAC
45	551	CGCACAAAGT	CGATCAGCCG	TTTCCAGGTA	AGGCGCGTTG	GGAAAAAATC
	601	CAACAGGCGG	GTTCGGTCAG	GGATNNGGAA	AGCATATGGC	CGGTATTGTA
	651	CGCGCANGGC	TGGCTGTCCG	CAGNNACGCA	CAACGGGCGC	GATTACGCCT
	701	TGTTTTTCCG	TCAGCCGGTT	CCCAAAGGCG	TGGCAGAGGA	TGCCGTCTTA
	751	ATCGAAAAGG	CAAGGGCGNA	ANANNNTNAG	TTGAGTTACA	GCAAAAAGG
50	801	TTTGACAGAC	TTTTTCCTNG	CAACCTTGCT	GATTGCCTCN	CTGCTGTCTA
	851	TTTTTCTTGC	ACTGGTCATG	GCACTGTATT	TCGCCCCGCC	TTTCGTGCAA
	901	CCCGTCTTAT	CGCTTGCCGA	GGGGGCGAAG	GCGGTGGCGC	AAGGCGGATT
	951	CAGCCAGACG	CGCCCCGTGT	TGCGCAACGA	CGAGTTCGGA	CGCTTGACCA

-236-

5
10
15
20
25

```

1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGACATTATC TCGAATGCGT
1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCTTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301 ACAAAACCGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACNGCAACG GCGTGGTAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
1601 AACAAAGTGGC GGCATTAATA GAAATGGTCG AGGCATTCCG CAATTACNCG
1651 CGTTCCTCCT CGNCTCAATT GGAAAATCAG GATTTGAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGGTTT GCGGCGGAAC
1751 TTGCCGGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TGTGCCCCGA GTCAGGGTAA AATCGGAAGC GGGGCAGGAC GGACGGATTG
1901 TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCAGGGA AATGCTGCAC
1951 AATGCCTTCG AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG
2001 ACTGCCCCGT GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC
2051 TGAGCAATCA GGATGCGGGC GCGCGTNTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G

```

25 This encodes a protein having amino acid sequence [<SEQ ID 254>] (SEQ ID NO: 254):

30
35
40

```

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QQAGSVRDxE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301 PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEEA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401 PLWGSSRHGW HGVSAAQSSL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL
35 451 LGKATVLPED NXNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDX DAQILTRSTD TIIKQVAALK EMVEAFRNYX
551 RSPSXQLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLM MAADTTAMRQ
601 VLHNIKFNAE EAAEADVPE VRVKSEAGQD GRIVLTVCDN GKGFGREMLH
651 NAFEPYVTDK PAGTGLXLPV VKKIIIEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*

```

ORF64a (SEQ ID NO: 254) and ORF64-1 (SEQ ID NO: 252) show 96.6% identity in 706 aa overlap:

45
50

```

              10      20      30      40      50      60
orf64a.pep  MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
              |||||
orf64-1      MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
              10      20      30      40      50      60

              70      80      90     100     110     120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              |||||
orf64-1      DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              70      80      90     100     110     120

```

		130	140	150	160	170	180
	orf64a.pep	SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK					
5	orf64-1	SKSALNLAADNALGNAVVPQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf64a.pep	SINPHKLDQPPFGKARWEKIQQAGSVRDYESIGGVLYAXGWLSAXTHNGRDYALFFRQPV					
10	orf64-1	SINPHKLDQPPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf64a.pep	PKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE					
15	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf64a.pep	PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA					
20	orf64-1	PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf64a.pep	RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMLPTPLWGSSRHGWHGVSAQQSLL					
25	orf64-1	RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMLPTPLWGSSRHGWHGVSAQQSLL					
		370	380	390	400	410	420
		430	440	450	460	470	480
	orf64a.pep	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQK					
30	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK					
		430	440	450	460	470	480
		490	500	510	520	530	540
	orf64a.pep	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTRSTDITIKQVAALK					
35	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTRSTDITIVKQVAALK					
		490	500	510	520	530	540
		550	560	570	580	590	600
	orf64a.pep	EMVEAFRNYXRSPSXQLENQDLNALIGDVLALYEAGPCRFAAELAGEPLMMAADTTAMRQ					
40	orf64-1	EMVEAFRNYARSPSLKLENQDLNALIGDVLALYEAGPCRFAAELAGEPLTVAADTTAMRQ					
		550	560	570	580	590	600
		610	620	630	640	650	660
	orf64a.pep	VLHNI FKNAAEAAEEADVPEVRVKSEAGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK					
45	orf64-1	VLHNI FKNAAEAAEEADVPEVRVKSETQDQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK					
		610	620	630	640	650	660
		670	680	690	700		
	orf64a.pep	PAGTGLXLPVVKKIIIEHGGXISLSNQDAGGAXVRIILPKTVETYAX					
50	orf64-1	PAGTGLGLPVVKKIIIEHGGRI SLSNQDAGGACVRIILPKTVKTYAX					
		670	680	690	700		

5
 10
 15
 20
 25
 30
 35
 40

201 CGGTTTCGCAG ATTGCCAAAC GCCTTTCCGG GATGTTACAG CTGGTCGCCC
 251 TACTGCCCCG CTTGTTCCTG TTCGGCATT TCCGCGCAGTT TATCAACGGC
 301 ACGATTAATT CGTGGTTCGG CAACGACACC CACGAAGCCC TCGAACGCAG
 351 CCTTAATTTG AGCAAGTCCG CACTGGATT TGGCGCAGAC AATGCCGTCA
 401 GCAACGCCGT TCCCGTACAG ATAGACCTCA TCGGCACCGC CTCCCTGTCTG
 451 GGCAATATGG GCAGTGTGCT GGAACACTAC GCCGGCAGCG GTTTTGCCCA
 501 GCTTGCCCTG TACAATGCCG CAAGCGGGAA AATCGAAAAA AGCATCAATC
 551 CGCACCAATT CGACAGCCG CTTCCCGACA AAGAACATTG GGAACAGATT
 601 CAGCAGACCG GTTCGGTTCTG GAGTTTGGA AGCATAGGCG GCGTATTGTA
 651 CGCGCAGGGA TGGTGTCTGG CAGGTACGCA CAACGGGCGC GATTACGCGC
 701 TGTTCCTCCG CCAGCCGATT CCCGAAAATG TGGCACAGGA TGCCGTTCTG
 751 ATTGAAAAGG CGCGGGCGAA ATATGCCGAA TTGAGTTACA GAAAAAAGG
 801 TTTGCAGACC TTTTCTCTGG TAACCCTGCT GATGCTCTCG CTGCTGTCTGA
 851 TTTTCTTGC GCTGTAAATG GCACTGTATT TTGCCCCTCG TTTCGTCTGAA
 901 CCCATTCTGT CGCTTGCCGA GGGCGCAAAG GCGGTGGCGC AGGGTGATT
 951 CAGCCAGACG CGCCCCGTAT TGCGCAACGA CGAGTTCGGA CGTTTGACCA
 1001 AGCTGTTCAA CCATATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
 1051 GAACGCAACC GCCGGCGCGA GGAAGCCGCC CGTCACTACC TCGAGTGCCT
 1101 GTTGATGGG TTGACTACCG GTGTGGTGGT GTTTGACGAA AAAGGCCGTT
 1151 TGAAAACCTT CAACAAGGCG GCGGAACAGA TTTTGGGGAT GCCGCTCGCC
 1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
 1251 GTCCCTGCTT GCCGAAGTGT Ttgcgcgcat CGGTGCGGCG GCAGGTACGG
 1301 ACAAAACCGT CCAGGTGGA TATGCGCGC CGGACGATGC CAAAATCCTG
 1351 CTGGGCAAGG CGACGTATT GCCCGAAGAC AACGGCAACG GCGTGGTGAT
 1401 GGTGATTGAC GACATCACCG TGCTGATACG CGCGCAAAAA GAAGCCGCGT
 1451 GGGGTGAAGT GCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCACG
 1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
 1551 GGACGATCAG GACGCGCAAA TCCTGACGCG TtcgACCGAC ACCATCATCA
 1601 AACAGgtggc gGCGTTAAAA GAAATGGTCG AGGCATTCCG CAATTACGCG
 1651 CGGCCCCCTT CGCTCAAACT GGAAAATCAG GATTGAACG CCTTAATCCG
 1701 CGATGTTTTG GCCCTGTACG AAGCCGGCCC GTGCCGTTT GAGGCGGAAC
 1751 TTGCCGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
 1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
 1851 TATGCCCGAA GTCAGGGTAA AATCGGAAAC GGGGCAGGAC GGACGGATTG
 1901 TCCTGACGGT TTGCGACAAC GGCAAGGGAT TCGGCAAGGA AATGCTGCAC
 1951 AATGCTTTCG AGCCGTATGT GACGGATAAG CCGGCGGGA CGGGACTGGG
 2001 TCTGCTGTA GTGAAAAAAA TCATTGGAGA ACACGGCGGC CGCATCAGCC
 2051 TGAGCAATCA GGATGCGGGT GGGGCGTGTG TCAGAATCAT CTTGCCAAAA
 2101 ACGGTAGAAA CTTATGCGTA G

This corresponds to the amino acid sequence [<SEQ ID 258; ORF64ng-1>] (SEQ ID NO: 258;

ORF64ng-1):

45
 50
 55

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV
 51 LARYVILLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING
 101 TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVVPQ IDLIGTASLS
 151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQP LPDKEHWEQI
 201 QQTGSVRSLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPI PENVAQDAVL
 251 IEKARAKYAE LSYSKKGLQT FFLVTLIAS LLSIFLALVM ALYFARRFVE
 301 PILSLAEGAK AVAQGDFSQT RPVLRNDEFQ RLTKLFNHMT EQLSIAKEAD
 351 ERNRRREEAA RHYLECVDLG LTTGVVVFDE KGRLKTFNKA AEQILGMPLA
 401 PLWGSSRHGW HGVSAQQSLL AEFVAAIGAA AGTDKPVQVE YAAPDDAKIL
 451 LGKATVLPED NGNGVMVID DITVLIRAQK EAAWGEVAKR LAHEIRNPLT
 501 PIQLSAERLA WKLGGKLDQ DAQILTRSTD TIIKQVAALK EMVEAFRNYA
 551 RAPSLKLENQ DLNALIGDVL ALYEAGPCRF EAELAGEPLM MAADTTAMRQ
 601 VLHNIFKNAA EAAEADMPE VRVKSETGQD GRIVLTVCDN GKGFGEMLH
 651 NAFEPYVTDK PAGTGLGLPV VKKIIGEHGG RISLSNQDAG GACVRIILPK
 701 TVETYA*

ORF64ng-1 (SEQ ID NO: 258) and ORF64-1 (SEQ ID NO: 252) show 93.8% identity in 706 aa overlap:

5	orf64ng-1.pep	10 20 30 40 50 60	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAMLLLVLSAVLARYVILLK
	orf64-1	10 20 30 40 50 60	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLK
10	orf64ng-1.pep	70 80 90 100 110 120	DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
	orf64-1	70 80 90 100 110 120	DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
15	orf64ng-1.pep	130 140 150 160 170 180	SKSALDLAADNAVSNAVVPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
	orf64-1	130 140 150 160 170 180	SKSALNLAADNALGNVAVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK
20	orf64ng-1.pep	190 200 210 220 230 240	SINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQPI
	orf64-1	190 200 210 220 230 240	SINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV
25	orf64ng-1.pep	250 260 270 280 290 300	PENVAQDAVLIEKARAKYAELSYSKKGLQTFVLVTLIASLLSIFLALVMALYFARRFVE
	orf64-1	250 260 270 280 290 300	PKGVAEDAVLIEKARAKYAELSYSKKGLQTFVLATLLIASLLSIFLALVMALYFARRFVE
30	orf64ng-1.pep	310 320 330 340 350 360	PILSLAEGAKAVAQGDQFSQTRPVLNRNDEFGRCLKLFNHNMTQELSIAKEADERNRREEAA
	orf64-1	310 320 330 340 350 360	PVLSLAEGAKAVAQGDQFSQTRPVLNRNDEFGRCLKLFNHNMTQELSIAKEADERNRREEAA
35	orf64ng-1.pep	370 380 390 400 410 420	RHYLECVDGLTTGVVVVFDEKGRCLKTFNKAAEQILGMPLAPLWGSSRHGWHGVSAQQSLL
	orf64-1	370 380 390 400 410 420	RHYLECVLGLTTGVVVVFDEQGLCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSLL
40	orf64ng-1.pep	430 440 450 460 470 480	AEVFAAIGAAAGTDKPVQVEYAAPDDAKILLGKATVLPEDNGNGVVMIDDITVLIRAQK
	orf64-1	430 440 450 460 470 480	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMIDDITVLIHAQK
45	orf64ng-1.pep	490 500 510 520 530 540	EAAWGEVAKRLAHEIRNPLTPQLSAERLAWKLGKGLDDQDAQILTRSTDITIKQVAALK
	orf64-1	490 500 510 520 530 540	EAAWGEVAKRLAHEIRNPLTPQLSAERLAWKLGKGLDEQDAQILTRSTDITIVKQVAALK

		550	560	570	580	590	600
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLMMAADTTAMRQ					
5	orf64-1	EMVEAFRNYARSPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLTVAADTTAMRQ					
		550	560	570	580	590	600
	orf64ng-1.pep	VLHNIFKNAAEAAEEADMPEVRVKSETGQDGRIVLTVCDNGKGFGKEMLHNAFEPYVTDK					
10	orf64-1	VLHNIFKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK					
		610	620	630	640	650	660
	orf64ng-1.pep	PAGTGLGLPVVKKIIEHGGRISLSNQDAGGACVRIILPKTVETYAX					
15	orf64-1	PAGTGLGLPVVKKIIEEHGGRISLSNQDAGGACVRIILPKTVKTYAX					
		670	680	690	700		

Furthermore, ORF64ng-1 (SEQ ID NO: 258) shows significant homology to a protein (SEQ ID NO: 1129) from *A.caulinodans*:

20	sp Q04850 NTRY_AZOCA NITROGEN REGULATION PROTEIN NTRY)gi 77479 pir S18624 ntrY protein - Azorhizobium caulinodans)gi 38737 (X63841) NtrY gene product [Azorhizobium caulinodans] Length = 771 Score = 218 bits (550), Expect = 7e-56 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)
25	Query: 7 IAAICAVVLLYGLTAATGSTSSLADYFWWIXXXXXXXXXXXXXXXXXXRYVILLKDRRNGV 66 I+A+ ++L GLT + + + R + + K R G Sbjct: 35 ISALATFLILMGLTPVVPHTQVVIS---VLLVNAAVLILSAMVGREIWRIAKARARGR 90
30	Query: 67 FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLKSALD 126 +++ R+ G+F +V+V+P + + +++ ++ ++ WF T E + S+++++ + Sbjct: 91 AAARLHIRIVGLFAVVSVPAILVAVVASLTLDRLDRWFSMRTQEIVASSVSVAQTYVR 150
35	Query: 127 LAADNAVSNAVPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP 184 A N + + + DL S+ Y G S F Q+ AA + ++ Sbjct: 151 EHALNIRGDILAMSADLTRLKSV-----YEGDRSRFNQILTAQAALRNLPGLMI 200
40	Query: 185 HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA----- 233 + D + ++ + I + V + +IG Q + N DY Sbjct: 201 RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQPVYLP--NDADYVAADVPLKDYDD 256
45	Query: 234 --LFFRQPIPENVAQDAVLIIEKARAKYAELSYSKKGLQTFVLVTXXXXXXXXXXXXXVMA 291 L+ + I V ++ A Y L + G+Q F + + Sbjct: 257 LYLYVARLIDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWL 316
	Query: 292 LYFARRFVEPILSLAEGAKAVAQGDQSQTRPVLRND-EFGRLTKLFNMTEQLSIXXXXX 350 L F++ V PI L A VA+G+ P+ R + + L + FN MT +L Sbjct: 317 LNFQSKWLVAPIRRLMSAADHVAEGNLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI 376
	Query: 351 XXXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRKTFNKAAEQILGMLPLAPLWGSSRHGW 410 + E VL G+ GV+ D + R+ N++AE++LG L+ + RH Sbjct: 377 LTARDQIDSRRRFTAVLSGVGAGVIGLDSQERITILNRSERLLG--LSEVEALHRHLA 434
	Query: 411 HGVSAQQSLLAEVFXXXXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM 467 V LL E + VQ D + + V E + +G V+

Sbjct: 435 EVVPETAGLLEEA-----EHARQRSVQGNITLTRDGRERVFAVRVTTEQSPEAEHGWVV 488

Query: 468 VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTR 527
 +DDIT LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL K G + QD +I +

Sbjct: 489 TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFGRHV-TQDREIFDQ 547

5 Query: 528 STDITIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGE 587
 TDTII+QV + MV+ F ++AR P +++QD++ +I + L G +

Sbjct: 548 CTDITIRQVGDIGRMVDEFSSFARMKPKVVDSDMSEIIRQTVFLMRVGHPEVVFDSVP 607

Query: 588 PLMAA-DTTAMRQVLHNIFKNXXXXXXXXDMPEVRVK-----SETGQDGRIVLTVCD 639
 P M A D + Q L NI KN P+VR + + G+D +V+ + D

10 Sbjct: 608 PAMPARFDRRLVSQALTNILKNAAEAIEAVP-PDVRGQGRIRVSANRVGED--LVIDIID 664

Query: 640 NGKGFGKEMLHNAFEPYVTDKPA GTGLGLPVVKKIIGE HGGRI SLNQDAG-GACVRIIL 698
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L

Sbjct: 665 NGTGLPQESRNRLLPEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined)
 15 and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is
 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 31

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 259] (SEQ ID
 20 NO: 259):

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
 51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
 101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
 151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
 25 201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCCTT TTGCTTTCCCT
 251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
 301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
 351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTGTATTTC AACAAATTAC
 401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG
 30 451 CACGCGTTGG ATACG...

This corresponds to the amino acid sequence [SEQ ID 260; ORF66] (SEQ ID NO: 260;
ORF66):

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
 35 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
 101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPNASTVIG
 151 HALDT...

Further work revealed the complete nucleotide sequence [SEQ ID 261] (SEQ ID NO: 261):

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
 51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
 101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
 151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
 5 201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTCCT
 251 ACGTCTTTTC CGTTTGTTC CACAACGGCA GTTGGACAGG CTGGGGCGCG
 301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
 351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATT CACAAATTAC
 401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC
 10 451 AACGCCTGG ATACGCTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
 501 CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
 551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCCG CTACGGCGTG
 601 ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC
 651 GCAAGACCGC CCCGCCCTT CGCTGCAAAA TCCGTAA

This corresponds to the amino acid sequence [<SEQ ID 262; ORF66-1>] (SEQ ID NO: 262; ORF66-1):

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
 20 101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPTASTVIG
 151 NALDTLVFFA VAFYASSDGF MAANWQGIAP VDYLFLKTV C TLFFLPAYGV
 201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*

Computer analysis of this amino acid sequence gave the following results:

25 Homology with the hypothetical protein o221 (SEQ ID NO: 1130) of *E. coli* (accession number P37619)

ORF66 (SEQ ID NO: 260) and o221 protein (SEQ ID NO: 1130) show 67% aa identity in 155aa overlap:

30 orf66 1 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
 M F+ Q+ KALF L LFH+L+I +SNYL VQ P I G HTTWGAFSFPFIFLATDLTV
 o221 1 MNVFSQTQRYKALFWLSLFLHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60
 orf66 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
 RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
 o221 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120
 35 orf66 121 IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155
 +GQILD+ VFN+LR+ + WW+AP AST+ G+ DT
 o221 121 LGQILDVHVFNRLRQSRWWLAPTASTLFGNVSDT 155

Homology with a predicted ORF from *N. meningitidis* (strain A)

40 ORF66 (SEQ ID NO: 260) shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) (SEQ ID NO: 264) from strain A of *N. meningitidis*:

-244-

5 orf66.pep MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
 orf66a MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV

10 orf66.pep RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
 orf66a RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA

15 orf66.pep IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT
 orf66a LGQILDIFVFNKLRLKAWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAP

orf66a VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX

The complete length ORF66a nucleotide sequence [<SEQ ID 263>] (SEQ ID NO: 263) is:

20 1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCTGGCTGGT
 51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
 101 CCTTCCAAAT TTCCGCGATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
 151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
 201 GGCACGGCGG ATTATCTTTT GGGTCATGTT CCCC GCCCTT TTGCTTTCTT
 25 251 ACGTCTTTTC CGTTTGTTC CACAACGGCA GTTGACGGG CTGGGCGCG
 301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTGC
 351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTTGTGTTT AACAAATTAC
 401 GCCGTCTGAA AGCGTGGTGG GTTGCCCCGA CTGCATCAAC CGTCATCGGC
 451 AACGCCTTAG ATACGTTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
 30 501 CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
 551 TGTTCAAACCT CACCGTCTGC GGTCTGTTTT TCCTGCCCCG CTACGGCGTG
 601 ATTCTGAATC TGCTGACGAA AAAACTGACG ACCCTGCAAA CCAAACAGGC
 651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

35 This encodes a protein having amino acid sequence [<SEQ ID 264>] (SEQ ID NO: 264):

40 1 MYAFTAAQQQ KALFWLVLFH ILIIAASNYL VQFPFQISGI HTTWGAFSFP
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
 101 LSEFNTFVGR IALASFAAYA LGQILDIFVF NKLRLKAWW VAPTASTVIG
 151 NALDTLVFFA VAFYASSDGF MAANWQGIAP VDYLFKLTV CGLFFLPAYGV
 201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*

ORF66a (SEQ ID NO: 264) and ORF66-1 (SEQ ID NO: 262) show 97.8% identity in 228 aa overlap:

45 orf66a.pep MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
 orf66-1 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV

-245-

	orf66a.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	
	orf66-1	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	
		70 80 90 100 110 120	
5	orf66a.pep	LGQILDIFVFNKLRLKAWVWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF	
	orf66-1	IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF	
		130 140 150 160 170 180	
10	orf66a.pep	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX	
	orf66-1	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX	
		190 200 210 220 229	

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 (SEQ ID NO: 260) shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) (SEQ ID NO: 266) from *N. gonorrhoeae*:

20	orf66.pep	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66ng	MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	120
	orf66ng	RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA	120
25	orf66.pep	IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT	155
	orf66ng	LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAF	180

The complete length ORF66ng nucleotide sequence [<SEQ ID 265>] (SEQ ID NO: 265) is:

30	1	ATGTACGCAT	TGACCGCCGC	ACAGCAACAG	AAGGCACTCT	TCCGGCTGGT
	51	GCTTTTCCAT	ATCCTCATCA	TCGCCGCCAG	CAACTATCTG	GTGCAGTTCC
	101	CCTTCCGGAT	TTTCGGCATC	CACACCACTT	GGGGCGCGTT	TTCCTTTCCC
	151	TTCATCTTCC	TCGCCACCGA	CCTGACCGTC	CGCATTTTCG	GTTTCGCACTT
35	201	GGCGCGGCGG	ATTATCTTTT	GGGTGATGTT	CCCCGCCCTT	ttgCTTcat
	251	acGTCTTTTC	CGTTTTGTTC	CACAACGGCA	GTTGGACGGG	CTTGGGCGCG
	301	ctgTCCCAAT	TCAACACCTT	TGTCGGACGC	ATCGCGCTGG	CAAGTTTTCG
	351	CGCCTACGCG	CTCGGACAAA	TCCTTGATAT	TTTCGTATTC	GACAAATTAC
40	401	GCCGTCTGAA	AGCGTGGTGG	ATTGCCCCGG	CCGCATCAAC	CGTCATCGGC
	451	AATGCACTGG	ACACGTTAGT	ATTTTTTGCC	GTTGCCTTTT	ACGCAAGCAG
	501	CGATGAATTT	ATGGCGGCAA	ACTGGCAGGG	CATCGCTTTT	GTGATTAC
	551	TGTTCAAAC	TACCGTCTGC	ACCCTCTTCT	TCCTGCCCGC	CTACGGCGTG
	601	ATACTGAATC	TGCTGACGAA	AAAAC	TGACG	GCCCTGCAAA
	651	GCAAGACCGC	CCCGTGCCCT	CGCTGCAAAA	TCCGTAA	

This encodes a protein having amino acid sequence [<SEQ ID 266>] (SEQ ID NO: 266):

45	1	MYALTAAQQQ	KALFRLVLFH	ILIIAASNYL	VQFPFRIFGI	HTTWGAFSFP
----	---	------------	------------	------------	------------	------------

51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL SLSYVFSVLF HNGSWTGLGA
 101 PSQNTFVGR IALASFAAYA LGQILDIFVF DKLRRLKAWW IAPAASTVIG
 151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLFKLTVC TLFFLPAYGV
 201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

An alternative annotated sequence is:

1 MYALTAAQQQ KALFRLVLFH ILIIAASNYL VQPPFRIFGI HTTWGAFSFP
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
 101 LSQNTFVGR IALASFAAYA LGQILDIFVF DKLRRLKAWW IAPAASTVIG
 151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLFKLTVC TLFFLPAYGV
 201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

ORF66ng (SEQ ID NO: 266) and ORF66-1 (SEQ ID NO: 262) show 96.1% identity in 228 aa overlap:

orf66-1.pep MYAFTAAQQQKALFRLVLFHILIIAASNYLVQPPFQIFGIHTTWGAFSFPFIFLATDLTV 60
 |||:|||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
 orf66ng MYALTAAQQQKALFRLVLFHILIIAASNYLVQPPFRIFGIHTTWGAFSFPFIFLATDLTV 60
 orf66-1.pep RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
 |||:|||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
 orf66ng RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
 orf66-1.pep IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF 180
 :|||||||||:|||||||||:|||||||||:|||||||||:|||||||||:|||||||||
 orf66ng LGQILDIFVFDKLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
 orf66-1.pep VDYLFKLTVC TLFFLPAYGVILNLLTKKLT TLQTKQAQDRPVPSLQNPX 229
 |||:|||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
 orf66ng VDYLFKLTVC TLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX 229

Furthermore, ORF66ng (SEQ ID NO: 266) shows significant homology with an *E.coli* ORF (SEQ ID NO: 1130):

sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC REGION (O221)
)gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli)gi|466607 (U00039) No definition line found [Escherichia coli])gi|1789882 (AE000423) hypothetical 25.3 kD protein in ftsY-nika intergenic region [Escherichia coli]
 Length = 221
 Score = 273 bits (692), Expect = 5e-73
 Identities = 132/203 (65%), Positives = 155/203 (76%)
 Query: 1 MYALTAAQQQKALFRLVLFHILIIAASNYLVQPPFRIFGIHTTWGAFSFPFIFLATDLTV 60
 M + Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
 Sbjct: 1 MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60
 Query: 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
 RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
 Sbjct: 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGWSQGFGALAHFNLFVARIATASFMAYA 120
 Query: 121 LGQILDIFVFDKLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
 LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA

Sbjct: 121 LGQILDVHVFNRLRQSRWWLAPTASTLFGNVSDTLAFFFIWFWRSPDAFMAEHWMEIAL 180

Query: 181 VDYLFLKLTVCFLFFLPAYGVILN 203
VDY FK+ + +FFLP YGV+LN

Sbjct: 181 VDYCFKVLISIVFFLPMYGVLLN 203

- 5 Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 32

- 10 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 267>] (SEQ ID NO: 267):

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAyGCA GTmwrAATAT
101 CTGAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
15 151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201 TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC
301 CGCTTAGgCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC
351 CcTTTtagCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
20 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGTAAA AGGCTACGAA
451 TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA
501 TGGCTGCTAC GCGCTTGAT..

- This corresponds to the amino acid sequence [<SEQ ID 268; ORF72>] (SEQ ID NO: 268; ORF72):

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAkIHkF
51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAgVGKLA
101 RLGAkFSTRA VPYVGtALLA HDVYETfKED IQARGYQYDP ETDKfVKGyE
30 151 YSNCLWYEDK RRINRTYGCY GVD..

Further work revealed the complete nucleotide sequence [<SEQ ID 269>] (SEQ ID NO: 269):

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
35 101 CTGAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351 CCTTTtagCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
40 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC
451 TAA

This corresponds to the amino acid sequence [<SEQ ID 270; ORF72-1>] (SEQ ID NO: 270; ORF72-1):

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVSG
151 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF72 (SEQ ID NO: 268) shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) (SEQ ID NO: 272) from strain A of *N. meningitidis*:

```

15 orf72.pep      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||||
orf72a      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS

20 orf72.pep      70      80      90     100     110     120
    DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPYVGTALLA
    |||||
orf72a      70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPYVGTALLA

25 orf72.pep      130     140     150     160     170
    HDVYETFKED IQARGYQYDP ETDKFKVKG YEYSNCLWY EDKRRINRTY GCGYGVGD
    |||||
orf72a      130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVSGX

```

The complete length ORF72a nucleotide sequence [<SEQ ID 271>] (SEQ ID NO: 271) is:

```

30 1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
    51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
    101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
    151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
    201 TTTAACACAC ATCCCTACGG GCGCAAAGC CCGAATCAAC GCCAAAATAA
    251 CCGCCAGCGT ATCCCGCGCC GCGTATTGG CGGGGTCGG CAAACTTGCC
35 301 CGCTTAGGCG CGAAATTCAG CACAAGGCG GTTCCCTATG TCGGAACAGC
    351 CCTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
    401 GAGGCTACCA ATACGACCCC GAAACGACA AATTTCACAA GGTCTCAGGC
    451 TAA

```

- 40 This encodes a protein having amino acid sequence [<SEQ ID 272>] (SEQ ID NO: 272):

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVSG

```


151 ★

ORF72a (SEQ ID NO: 272) and ORF72-1 (SEQ ID NO: 270) show 100.0% identity in 150 aa overlap:

5		10	20	30	40	50	60
	orf72a.pep	MVIKYTNLNF	AKLSIIAILM	MYSF	EANANAVK	ISETVS	VDTGQGAKIH
	orf72-1	MVIKYTNLNF	AKLSIIAILM	MYSF	EANANAVK	ISETVS	VDTGQGAKIH
		10	20	30	40	50	60
10		70	80	90	100	110	120
	orf72a.pep	DLIKTV	DLTHIPTG	AKARINAK	ITASVSR	AGVL	AGVGKL
	orf72-1	DLIKTV	DLTHIPTG	AKARINAK	ITASVSR	AGVL	AGVGKL
		70	80	90	100	110	120
15		130	140	150			
	orf72a.pep	HDVYET	FKEDIQ	ARGYQY	DPETDK	FAKVSGX	
	orf72-1	HDVYET	FKEDIQ	ARGYQY	DPETDK	FAKVSGX	
		130	140	150			

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF72 (SEQ ID NO: 268) shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) (SEQ ID NO: 274) from *N. gonorrhoeae*:

25	orf72.pep	MVIKYTNLNFAKLSIIAILMMYSFEANANAVXISETVSVDTGQGAKIHKFVPKNSKTYSS	60
	orf72ng	MVTKHTNLNFAKLSIIAILMMYSFEANANAVKISETLSVDTGQGAKVHKFVPKSSNIYSS	60
	orf72.pep	DLIKTVDLTHXPTGAKARINAKITASVSRAGVLAVGVGKLARLGAKFSTRAPVYVGTALLA	120
	orf72ng	DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVGKLVROGAKFGTRAPVYVGTALLA	120
30	orf72.pep	HDVYETFKEDIQARGYQYDPETDKFVKGYEYSNCLWYEDKRRINRITYGCVGD	173
	orf72ng	HDVYETFKEDIQARGCRYDPETDKFVKGYEYANCLWYEDERRINRITYGCVGDSSIMRLM	180

An ORF72ng nucleotide sequence [[SEQ ID 273](#)] ([SEQ ID NO: 273](#)) was predicted to encode a protein having amino acid sequence [[SEQ ID 274](#)] ([SEQ ID NO: 274](#)):

35	1	MVTKHTNLNF	AKLSIIAILM	MYSFEANANA	VKISETLSVD	TGQGAHVHKF
	51	VPKSSNIYSS	DLTKAVDLTH	IPTGAKARIN	AKITASVSRA	GVLSGVGKLV
	101	RQGAKEGTRA	VPYVGTALLA	HDVYETFKED	IQARGCRYDP	ETDKFVKGYE
	151	YANCLWYEDE	RRINRTYGCY	GVDSSIMRLM	PDRSRFPEVK	QLMESQMYRL
40	201	ARPFWNWRKE	ELNKLSSLDW	NNFVLNRCTF	DWNGGGCAVN	KGDDFRAGAS
	251	FSLGRNPKYK	EEMDAKKPEE	ILSLKVDADP	DKYIEATGYR	GYSEKVEVAP
	301	GTKVNMGPVT	DRNGNPVQVA	ATFGRDAQGN	TTADQVQVIP	PDLTPASAEA
	351	PHAQPLPEVS	PAENPANNPD	PDENPGTRPN	PEPDPDLNPD	ANPDTDGPQG
	401	TSPDSPAVPD	RPNGRHRKER	KEGEDGGLSC	DYFPEILACQ	EMGKPSDRMF

451 HDISIPQVTD DKTWSSHNFL PSNGVCPQPK TFHVFGQRQYR ASYEPLCVFA
501 EKIRFAVLLA FIIMSAFVVF GSLGGE*

After further analysis, the following gonococcal DNA sequence [<SEQ ID 275>] (SEQ ID NO: 275) was identified:

10

1	ATGGTCACAA	AACATACAAA	TTTGAATTTT	GCGAAATTGT	CGATAATTGC
51	AATTTTGTATG	ATGTATTTCGT	TTGAAGCGAA	TGCAAATGCA	GTAAAAATAT
101	CTGAAACTCT	TTCGGTTGAT	ACCGGACAAG	GCGCGAAAGT	TCATAAGTTC
151	GTTCTTAAAT	CAAGTAATAT	TTATTTCATCT	GATTTTAACAA	AAGCGGTAGA
201	TTTAACGCAT	ATCCCCACGG	GCGCAAAAGC	CCGAATCAAC	GCCAAAATAA
251	CCGCCAGCGT	ATCCCCGCGCC	GGCGTATTGT	CGGGGGTCTG	CAAACCTGTC
301	CGCCAAGGCG	CGAAATTCGG	CACAAGGGCG	GTTCCCTATG	TCGGAACAGC
351	CCTTTTAGCC	CACGACGTAT	ACGAAACTTT	CAAAGAAGAC	ATACAGGCAC
401	GAGGCTGCCG	ATACGATCCC	GAAACCGACA	AATTT	

This corresponds to the amino acid sequence [[SEQ ID 276; ORF72ng-1](#)] ([SEQ ID NO: 276; ORF72ng-1](#)):

20

1	MVTKHTNLNF	AKLSIIAILM	MYSFEANANA	VKISETLSVD	TGQGAHVHKF
51	VPKSSNIYSS	DLTKAVDLTH	IPTGAKARIN	AKITASVSRA	GVLSGVGKLV
101	ROGAKEFGTRA	VPYVGTALLA	HDVYETFKED	IQARGCRYDP	ETDKF

ORF72ng-1 (SEQ ID NO: 276) and ORF721-1 (SEQ ID NO: 270) show 89.7% identity in 145 aa overlap:

[illegible]

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 33

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 277>] (SEQ ID NO: 277):

```

5      1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
      51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
     101  TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
     151  GCTGACCGGT CTTTTATTGG CGGGCGCGGC AATGAGAAGC GCGGGAAGG
     201  TATCCGTTTA TCAGATGTTG TGGCCTATC.

```

10 This corresponds to the amino acid sequence [<SEQ ID 278; ORF73>] (SEQ ID NO: 278; ORF73):

```

      1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLRQTG
     51  LTGLLLAGAA MRSGGKVSIV QMLWPI..

```

15 Further work revealed the complete nucleotide sequence [<SEQ ID 279>] (SEQ ID NO: 279):

```

      1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
     51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
    101  TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCATACGGGG
    151  CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
    201  ATCCGTTTAT CAGATGTTGT GGCCTATCCG TTATACGGTG GCGGCTGTGT
    251  GTCTGATGAG TCCGGGATTC GTATCCTCGG TGTGGCGGT ATTGCTGCTG
    301  CTGCCGTTTA AGGGAGGGGC AGTGTTCAG GCAGGAGGTG CGGAAATTT
    351  TTTCAACATG AACCAATCGG GCAGAAAAGA GGGCTTTTCC CGCGATGACG
    401  ATATTATCGA GGGAGAATAT ACGTTGAAG AGCCTTACGG CGGCAATCGT
    451  TCCCAGAACG CCATCGAACA CAAAAAGAC GAATAA

```

This corresponds to the amino acid sequence [<SEQ ID 280; ORF73-1>] (SEQ ID NO: 280; ORF73-1):

```

30      1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLRHTG
     51  LSGLLLAGAA MRSGGRVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
    101  LPFKGGAVLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEPYGGNR
    151  SRNAIEHKKD E*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 (SEQ ID NO: 278) shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) (SEQ ID NO: 282) from strain A of *N. meningitidis*:

```

                                10      20      30      40      50      60
orf73.pep  MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMRLRQTGLTGLLLAGAA

```

```

|||||
orf73a  MRFFGIGFLVLLFLEIMSVVWVADWLGGGWTFLMAATFAAGVVMLRHTGLSGLLLAGAA
          10          20          30          40          50          60
          70
5  orf73.pep MRSGGKVSVYQMLWPI
          |||:|||||
orf73a  MRSGGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM

```

The complete length ORF73a nucleotide sequence [<SEQ ID 281>] (SEQ ID NO: 281) is:

10	1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
	51	GTCGATTGTG	TGGGTTGCCG	ATTGGTTGGG	CGGCGGTTGG	ACGCTGTTTC
	101	TAATGGCGGC	AACCTTTGCC	GCCGGCGTGG	TGATGCTCAG	GCATACGGGG
	151	CTGTCCGGTC	TTTTATTGGC	GGGCGCGGCA	ATGAGAAGCG	GCGGGAGGGT
	201	ATCCGTTTAT	CANATGTTGT	GGCNTATCCG	TTATACGGTG	GCGGCGGTGT
15	251	GTCNGATGAG	TCCGGGATTC	GTATCCTCGG	TGTNGGCGGT	ATTGCTGNTG
	301	CTNCCGTTTA	AGGGAGGTGC	AGTGTTGCAG	GCAGGAGGTG	CGGAAAATTT
	351	TTTCAACATG	AACCANTCGG	GCAGAAAAGA	NGGCNTTTC	CGCGATGACG
	401	ATATTATCGA	GGGGGAATAT	ACGGTTGAAG	ANCCTTACGG	CGGCANTCGT
	451	TTCCGAAACG	CCNTNGAACA	CAAAAAAGAC	GAATAA	

This encodes a protein having amino acid sequence [<SEQ ID 282>] (SEQ ID NO: 282):

25

```

      1 MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAATFA AGVVMLRHTG
     51 LSGLLLAGAA MRSGGRVSVY XMLWXIRYTV AAVCXMSPGF VSSVXAVLLX
    101 LPFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
    151 FRNAXEHKKD E*
```

ORF73a (SEQ ID NO: 282) and ORF73-1 (SEQ ID NO: 280) show 91.3% identity in 161 aa overlap

		10	20	30	40	50	60
30	orf73a.pep	MRFFGIGFLVLLFLEIMSI	VVWADWLGGGWTLFLMAATFAAGV	MLRHTGLSGLLLAGAA			
	orf73-1	MRFFGIGFLVLLFLEIMSI	VVWADWLGGGWTLFLMAAGFAAGV	MLRHTGLSGLLLAGAA			
		10	20	30	40	50	60
		70	80	90	100	110	120
35	orf73a.pep	MRS	GGRVSVYXMLWXIRYT	VA	AVCX	MSPGFVSSVXAVLLXL	LPFKGGAVLQAGGAENFFNM
	orf73-1	MRS	GGRVSVYQMLWP	IRYT	VA	AVCLMSPGFVSSVLAVLLLL	LPFKGGAVLQAGGAENFFNM
		70	80	90	100	110	120
		130	140	150	160		
40	orf73a.pep	NXSGRKXGXS	RDDDI	IEGEYTV	EXPYGGXRFRNAXE	HKKDEX	
	orf73-1	NQSGRK	EGFSRDDDI	IEGEYTV	EEPYGGRNR	RNAIEHKKDEX	
		130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 (SEQ ID NO: 278) shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) (SEQ ID NO: 284) from *N. gonorrhoeae*:

```

5  orf73.pep      MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAAGFAAGVLMRLQTGLTGLLLAGAA      60
    |||||
    orf73ng       MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAATFAAGVLMRLRHTGLSGLLLAGAA      60

    orf73.pep      MRSGGKVSQMLWPI                                          76
    ::|||
    orf73ng       VKSSGKVSQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM  120

```

10 The complete length ORF73ng nucleotide sequence [<SEQ ID 283>] (SEQ ID NO: 283) is:

```

1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAAATTAT
51  GTCGATTGTG TGGGTGCGCG ATTGGCTGGG CGGCGGTTGG AcgcTGTTTC
101 TAATGGCGGC AACCTTTGCC GCCGGTGTGC TGATGCTCAG GCATAcggGG
15  151 CTGTCCGGTC TTTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
    201 ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
    251 GTCTGatgag tCcgGATTc GTATCCTccg tgttggCGGT ATTGCTGCTG
    301 CTGCcgttta aggGaggGgc agtgttgca gaggagggtg cggaaaATTT
    351 TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
    401 atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
20  451 tcccgAAAcg ccatcgaaca cgaaaAagac gaataA

```

This encodes a protein having amino acid sequence [<SEQ ID 284>] (SEQ ID NO: 284):

```

25  1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAATFA AGVLMRLHTG
    51  LSGLLLAGAA VKSSGKVSQY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
    101 LPFGKGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGNNR
    151 SRNAIEHEKD E*

```

ORF73ng (SEQ ID NO: 284) and ORG73-1 (SEQ ID NO: 280) show 93.8% identity in 161 aa overlap

```

30  orf73-1.pep      10      20      30      40      50      60
    MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAAGFAAGVLMRLRHTGLSGLLLAGAA
    orf73ng         10      20      30      40      50      60
    MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTFLMAATFAAGVLMRLRHTGLSGLLLAGAA

35  orf73-1.pep      70      80      90      100     110     120
    MRSGGRVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM
    orf73ng         70      80      90      100     110     120
    VKSSGKVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM

40  orf73-1.pep      130     140     150     160
    NQSGRKEGFSRDDDIIEGEYTVVEEPYGGNRSRNAIEHKKDEX
    orf73ng         130     140     150     160
    NQSGRKEGFFHDDDIIEGEYTVVEKPDGNNRSRNAIEHEKDEX

45

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 34

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 285] (SEQ ID NO: 285):

```

1  ATGTTTGT TTTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
51  AGCCTCCGAC AGCGTCGTG GAGGGACATT ATACGTGGTT GCCACGCCCA
10  101 TCGGCAATTT GCGGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
151  GCG..... .GCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
201  CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
251  GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
301  GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
15  351 ACTCGCCCGC CGCGTGC GTG AGGCCGGGT TAAAGTCGTT CCCGTCGTGG
401  GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
451  GATTTTATT TCAACGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA
501  ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
551  CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
20  601 GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
651  CTTAAGCGGC ACGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
701  ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
751  AAACACGAAG GCTTGTCGGA GTCCGCGCAA AACATCATGA AAATCCTCAC
801  AGCCGAGCTG CCGACCAAAC AGGCGCGGGA GCTTGCTGCC AAAATCACGG
25  851 GCGAGGGAAA GAAAGCTTTG TACGAT..

```

This corresponds to the amino acid sequence [SEQ ID 286; ORF75] (SEQ ID NO: 286; ORF75):

```

1  MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
30  51  A...AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMV
101  AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGAXAVMA ALSVAGVEGS
151  DFYFNGFVPP KSGERRKLFA KWVRAAFPIV MFETPHRIGA ALADMAELFP
201  ERRMLMLAREI TKTFETFLSG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE
35  251  KHEGLSESAQ NIMKILTAEL PTKQAELAA KITGEGKKAL YD..

```

Further work revealed the complete nucleotide sequence [SEQ ID 287] (SEQ ID NO: 287):

```

1  ATGTTTCAGA AACATTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51  ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
40  101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151  CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAG GCAAATCGT
201  CAGTGTGCGC GAACACAACG AACGGCAGAT GCGGACAAAG ATTGTCGGCT
251  ATCTTTTACA CGGCATGGT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301  GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351  GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
45  401 GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTGTACCG

```

451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGGG
 501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
 551 CCGATATGGC GGAACGTGTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
 601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
 701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
 751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
 801 GGAGCTTGCT GCCAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
 851 TGGCTCTGTC TTGGAAAAAC AATAG

This corresponds to the amino acid sequence [<SEQ ID 288; ORF75-1>] (SEQ ID NO: 288; ORF75-1):

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQKGLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP
 101 AVCDPQAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLMLARE
 201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 (SEQ ID NO: 286) shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) (SEQ ID NO: 290) from strain A of *N. meningitidis*:

25	orf75.pep	10	20	30	40	50	60
		MFVFQTA	FXMFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KAXXXAEDTR
	orf75a		MFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDTR
			10	20	30	40	50
30	orf75.pep	70	80	90	100	110	120
		VTAQLLSAY	GIQKGLVSVR	EHNERQMA	DKIVGYLSDGMV	VAQVSDAGTP	PAVCDPQAKLAR
	orf75a	VTAQLLSAY	GIQKGLVSVR	EHNERQMA	DKIVGYLSDGMV	VAQVSDAGTP	PAVCDPQAKLAR
		60	70	80	90	100	110
35	orf75.pep	130	140	150	160	170	180
		RVREAGFKV	VPVVGAXAVMA	ALSVAGVEG	SDFYFNGFV	PPKSGERRKLF	AKWVRAAFPIV
	orf75a	RVREVGFKV	VPVVGASAVMA	ALSVAGVAG	SDFYFNGFV	PPKSGERRKLF	AKWVRVAFPVV
		120	130	140	150	160	170
40	orf75.pep	190	200	210	220	230	240
		MFETPHRIGA	ALADMAELF	PERRLMLARE	ITKTFETFLS	GTVGEIQTAL	SADGNQSRGEM
	orf75a	MFETPHRIGA	TLADMAELF	PERRLMLARE	ITKTFETFLS	GTVGEIQTAL	AADGNQSRGEM
		180	190	200	210	220	230
45	orf75.pep	250	260	270	280	290	
		VLVLYPAQ	DEKHEGLSESA	QNIMKILTAE	LPTKQAAELA	AKITGEGK	KALYD

orf75a VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNK
240 250 260 270 280 290

orf75a X

5 The complete length ORF75a nucleotide sequence [<SEQ ID 289>] (SEQ ID NO: 289) is:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTGGCGGAC ATTACCCTGC
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CGCGTTACCG CGCAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACCTCGT
201 CAGCGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT
10 ATCTTTTCTAGA CGGCATGGTT GTGGCACAGG TTCCGATGC GGGTACGCCG
251 ATCTTTTCTAGA CGGCATGGTT GTGGCACAGG TTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGTCGG
351 GTTTAAAGTT GTCCTTGTG TCGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GTGTGGCTGG TGTGGCGGGA TCCGATTTTT ATTCAACGG TTTTGTACCG
15 451 CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TCGGGGTGGC
501 GTTTCCCGTC GTGATGTTTG AAACGCCGCA CCGCATCGGG GCGACGCTTG
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
601 ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG
20 701 TGCTTTATCC GCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAAAATCA TGAATCTCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801 GGAGCTTGCC GCCAAATCA CGGCGGAGGG AAAAAAGCT TTGTACGATC
851 TGGCACTGTC TTGAAAAAC AAATGA

25 This encodes a protein having amino acid sequence [<SEQ ID 290>] (SEQ ID NO: 290):

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTQQLLSAY GIQKLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP
101 AVCDPGAKLA RRVREVGFKV VPVVGASAVM AALSVAGVAG SDFYFNGFVP
151 PKSGERRKLF AKWVRVAFPV VMFETPHRIG ATLDMAELF PERRMLLARE
30 201 ITKTFFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

ORF75a (SEQ ID NO: 290) and ORF75-1 (SEQ ID NO: 288) show 98.3% identity in 291 aa overlap:

35 orf75a.pep 10 20 30 40 50 60
orf75-1 MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQQLLSAY
10 20 30 40 50 60
40 orf75a.pep 70 80 90 100 110 120
orf75-1 GIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVGFKV
70 80 90 100 110 120
45 orf75a.pep 130 140 150 160 170 180
orf75-1 VPVVGASAVMAALSVAGVAGSDFYFNGFVPPKSGERRKLF AKWVRVAFPVVMFETPHRIG
130 140 150 160 170 180

-257-

		190	200	210	220	230	240
orf75a.pep	m	ATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQD					
orf75-1		ATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD					
		190	200	210	220	230	240
orf75a.pep		250	260	270	280	290	
orf75-1		EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX					
		250	260	270	280	290	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF75 (SEQ ID NO: 286) shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) (SEQ ID NO: 292) from *N. gonorrhoeae*:

orf75.pep	MFVFQTAFXMFQKHLQKASDSVVGGLTYVVPATPIGNLADITLRALAVLQKA---AEDTR	56
orf75ng	MSVFQTAFFMFQKHLQKASDSVVGGLTYVVPATPIGNLADITLRALAVLQKADIICAEDTR	60
orf75.pep	VTAQLLSAYGIQGKLVSREHNERQMA DKIVGYLS DGMVVAQVSDAGTPAVCDPGAKLAR	116
orf75ng	VTAQLLSAYGIQGR LVSREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
orf75.pep	RVREAGFKVVPVVGAXAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFKAWVRAAFPV	176
orf75ng	RVREAGFKVVPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKAWVRAAFPV	180
orf75.pep	MFETPHRIGAAALADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALSADGDQSRGEM	236
orf75ng	MFETPHRIGATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM	240
orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELA AKITGEGKKALYD	288
orf75ng	VLVLYPAQDEKHEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence [<SEQ ID 291>] (SEQ ID NO: 291) was predicted to encode a protein having amino acid sequence [<SEQ ID 292>] (SEQ ID NO: 292):

1	MSVFQTAFFM	FQKHLQKASD	SVVGGLTYVV	ATPIGNLADI	TLRALAVLQK
51	ADIICAEDTR	VTAQLLSAYG	IQGR LVSRE	HNERQMA DKV	IGFLSDGLVV
101	AQVSDAGTPA	VCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSVAGVAES
151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP
201	ERRMLLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
301	*				

After further analysis, the following gonococcal DNA sequence [<SEQ ID 293>] (SEQ ID NO: 293) was identified:

1 ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85

51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCAGAC ATTACCTGCG
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATTTGTGC CGAAGACACG
151 CCGCTTACTG CGCAGCTTTT GAGCGCGTAC GGCATTTCAGG GCAGGTTGGT
201 CAGTGTGCGC GAACACAACG AGCGGCAGAT GGCGGACAAG GTAATCGGTT
251 TCCTTTCAGA CGGCCTGGTT GTGGCGCAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GCGAAGCAGG
351 GTTCAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTAATG GCGGCGTTGA
401 GTGTGGCCGG TGTGGCGGAA TCCGATTTTT ATTTCACGG TTTTGTACCG
451 CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TGCGGGCGGC
501 ATTTCTGTGC GTCATGTTTG AAACGCCGCA CCGAATCGGG GCAACGCTTG
551 CCGATATGGC GGAATTGTTC CCCGAACGCC GTCTGATGCT GGCGCGCGAA
601 ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG GCGCGGACG GCAACCAATC GCGCGCGCAG ATGGTGTTGG
701 TGCTTTATCC GGCAGCAGAT GAAAAACACG AAGGCTTGTC CGAGTCTGCG
751 CAAATGCGA TGAATTCCT TGGCGCCGAG CTGCCGACCA AGCAGGCGGC
801 GGAGCTTGCC GCCAAGATTA CAGGTGAGGG CAAAAAGGCT TTGTACGATT
851 TGGCACTGTC GTGAAAAAC AAATGA

This corresponds to the amino acid sequence [<SEQ ID 294; ORF75ng-1>] (SEQ ID NO: 294;

20 ORF75ng-1):

25
30
35
40
45
50
55
60
65
70
75
80
85

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQGRVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRLMLARE
201 ITKTFTFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

ORF75ng-1 (SEQ ID NO: 294) and ORF75-1 (SEQ ID NO: 288) show 96.2% identity in 291 aa overlap:

30
35
40
45
50

orf75-1.pep
orf75ng-1
orf75-1.pep
orf75ng-1
orf75-1.pep
orf75ng-1
orf75-1.pep
orf75ng-1
orf75-1.pep
orf75ng-1

10 20 30 40 50 60
10 20 30 40 50 60
70 80 90 100 110 120
70 80 90 100 110 120
130 140 150 160 170 180
130 140 150 160 170 180
190 200 210 220 230 240
190 200 210 220 230 240
250 260 270 280 290

```

orf75-1.pep  EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX
              |||||:|||||
orf75ng-1    EKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNKX
              250      260      270      280      290

```

- 5 Furthermore, ORG75ng-1 (SEQ ID NO: 294) shows significant homology to a hypothetical *E.coli* protein (SEQ ID NO: 1131):

```

10 sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
   (F286)
   )gi|606086 (U18997) ORF f286 [Escherichia coli]
   )gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
   [Escherichia coli] Length = 286
   Score = 218 bits (550), Expect = 3e-56
   Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

15 Query: 4   KHLQKASDSVVGGLTVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
      K  Q A +S  G LY+V TPIGNLADIT RAL VLQ  D+I AEDTR T  LL  +GI
      Sbjct: 2   KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLLQHFGIN 59

      Query: 64  GRLVSVREHNERQMADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
      RL ++ +HNE+Q A+ ++  L +G  +A VSDAGTP + DPG  L R  REAG +VVP+
      Sbjct: 60  ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

20 Query: 124  VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFQKQVRAAFVVMFETPHRIGATL 183
      G  A + ALS AG+  F + GF+P KS  RR  ++ +E+ HR+  +L
      Sbjct: 120 PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAIKAEPRTLIFYESTHRLLDL 179

      Query: 184  ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
      D+  +  E R ++LARE+TKT+ET  VGE+  +  D N+  +GEMVL++  +
25 Sbjct: 180  EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

      Query: 243  HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286
      E  L  A  + +L AELP K+AA LAA+I G  K ALY  AL
      Sbjct: 239  EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

```

- Based on this analysis, including the presence of a putative transmembrane domain in the
 30 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 35

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 295>] (SEQ ID
 NO: 295):

```

35      1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
      51  TTTTGCAGCA GC.AAAGCAC CCGAAATCGA CCCGCTTTG .....
           //
      651  ..... .GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTCGAGAC
      701  AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAGA AACGGGTGTC
40      751  AAACCGTAA

```

This corresponds to the amino acid sequence [<SEQ ID 296; ORF76>] (SEQ ID NO: 296; ORF76):

```

1  MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
5  201 ..... ELVRNQLEQG LRQEKARLKI DALLEENGVK
    251 P*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 297>] (SEQ ID NO: 297):

```

10  1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
    51  TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
    101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
    151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC
    201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
    251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
    15  301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAGACGAGCT
    351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
    401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
    451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
    501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC
    20  551 AGTTTGCCGC GATGAATCGG GCGGACGTTA CCCGCGATCC GGTCAAATTG
    601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
    651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAGCAG GGTTCGAGAC
    701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGGAAGA AAACGGTGTC
    751 AAACCGTAA
    25

```

This corresponds to the amino acid sequence [<SEQ ID 298; ORF76-1>] (SEQ ID NO: 298; ORF76-1):

```

30  1  MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
    51  KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNRF KIAEASFYAE
    101 EYVRFLERSE TVSEDELHKF YEQQIRMIKL QVVSFATEEE ARQAQQLLLK
    151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
    201 GERYYLFLKS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENGV
    251 KP*

```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 (SEQ ID NO: 296) shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) (SEQ ID NO: 300) from strain A of *N. meningitidis*:

```

40  orf76.pep      10      20      30
    MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL
    |||||
    orf76a        10      20      30      40      50      60
    MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
    |||||
    //
    45  orf76.pep      70      80      90
    XELVRNQLEQGLRQEKARLKIDALLEENGVKPX
    |||||

```

orf76a DVTRDPVKLGERYLFLKLSVGNPDAPPFELVRNQLEQGLRQEKARLKIDAILEENGVKPX
200 210 220 230 240 250

The complete length ORF76a nucleotide sequence [<SEQ ID 299>] (SEQ ID NO: 299) is:

```
5      1  ATGAAACAGA  AAAAAACCGC  TGCCGCAGTT  ATTGCTGCAA  TGTTGGCAGG
      51  TTTTGCGGCA  GCCAAAGCAC  CCGAAATCGA  CCCGGCTTTG  GTGGATACGC
     101  TGGTGGCGCA  GATCATGCAG  CAGGCAGACC  GGCATGCGGA  GCAGTCCCAA
     151  AAACCGGACG  GGCAGGCAAT  CCGAAACGAT  GCCGTCCGTC  GGCTGCAAAAC
     201  TTTGGAAGTT  TTGAAAAACA  GGGCATTGAA  GGAAGGTTTG  GATAAGGATA
    10  251  AGGATGTCCA  AAACCGCTTT  AAAATCGCCG  AAGCGTCTTT  TTATGCCGAG
     301  GAGTACGTCC  GTTTTCTGGA  ACGTTCGGAA  ACGGTTTCCG  AAAGCGCACT
     351  GCGTCAGTTT  TATGAGCGGC  AAATCCGCAT  GATCAAATTG  CAGCAGGTCA
     401  GCTTCGCAAC  CGAAGAGGAG  GCGCGTCAGG  CGCAGCAGCT  CCTGCTCAAA
     451  GGGCTGTCTT  TTGAAGGGCT  GATGAAGCGT  TATCCGAACG  ACGAGCAGGC
    15  501  TTTTGACGGT  TTCATTATGG  CGCAGCAGCT  TCCCGAGCCG  CTGGCTTCGC
     551  AGTTTGACGC  GATGAATCGG  GCGGACGTTA  CCCGCGATCC  GGTCAAATTG
     601  GGCGAACGCT  ATTATCTGTT  CAAACTCAGC  GAGGTCGGGA  AAAACCCCGA
     651  CGCGCAGCCT  TTCGAGTTGG  TCAGAAACCA  GTTGGAACAA  GGTTTGAGAC
     701  AGGAAAAAGC  CCGCTTGAAA  ATCGATGCCA  TTTTGGAAGA  AAACGGTGTC
    20  751  AAACCGTAA
```

This encodes a protein having amino acid sequence [<SEQ ID 300>] (SEQ ID NO: 300):

```
25      1  MKQKKTAADV  IAAMLAGFAA  AKAPEIDPAL  VDTLVAQIMQ  QADRHAEQSQ
     51  KPDGQAIRND  AVRRLQTLEV  LKNRALKEGL  DKDKDVQNR  KIAEASFYAE
     101  EYVRFLELSE  TVSESALRQF  YERQIRMIKL  QQVSFATEEE  ARQAQQLLLK
     151  GLSFEGLMKR  YPNDEQAFDG  FIMAQQLPEP  LASQFAAMNR  GDVTRDPVKL
     201  GERYLFLKLS  EVGKNPDAP  FELVRNQLEQ  GLRQEKARLK  IDAILEENG
     251  KP*
```

30 ORF76a (SEQ ID NO: 300) and ORF76-1 (SEQ ID NO: 298) show 97.6% identity in 252 aa overlap:

```
35      10      20      30      40      50      60
orf76a.pep  MKQKKTAADVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
      |||
orf76-1     MKQKKTAADVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
      10      20      30      40      50      60

      70      80      90      100     110     120
orf76a.pep  AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLELSETVSESALRQF
      |||
orf76-1     AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLELSETVSEDELHKF
      70      80      90      100     110     120

      130     140     150     160     170     180
orf76a.pep  YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      |||
orf76-1     YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      130     140     150     160     170     180

      190     200     210     220     230     240
orf76a.pep  LASQFAAMNRGDVTRDPVKLGERYLFLKLSVGNPDAPPFELVRNQLEQGLRQEKARLK
      |||
orf76-1     LASQFAAMNRGDVTRDPVKLGERYLFLKLSVGNPDAPPFELVRNQLEQGLRQEKARLK
```

		190	200	210	220	230	240
		250					
	orf76a.pep	IDAILEENGVKPX					
5	orf76-1	:					
		IDAILEENGVKPX					
		250					

Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 (SEQ ID NO: 296) and a predicted ORF (ORF76.ng) (SEQ ID NO: 302) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

	orf76.pep	MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL	30
	orf76ng	MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDQGAIRND	60
15	orf76.pep	//	
		ELVRNQLEQGLRQEKARLKIDALLEENGVKP	251
	orf76ng	VTRNPVKLGERYYLFLKLGAVGKNPDAQPFELVRNQLEQGLRQEKARLKIDALLEENGVKP	251

The complete length ORF76ng nucleotide sequence [<SEQ ID 301>] (SEQ ID NO: 301) is:

20	1	ATGAAACAGA	AAAAGACCGC	TGCCGCAGTT	ATTGCTGCAA	TGTTGGCAGG
	51	TTTTCGCGCA	GCCAAAGCAC	CCGAAATCGA	CCCAGCTTTG	GTGGATACGC
	101	TGGTGGCGCA	GATCATGCAG	CAGGCAGACC	GGCATGCGGA	GCAGTCCCAA
	151	AGACCGGACG	GGCAGGCAAT	CCGAAACGAT	GCCGTCCGCC	GGCTGCAAAC
	201	TTTGGAAGTT	TTGAAAAACA	GGGCATTGAA	GGAAGGTTTG	GATAAGGATA
25	251	AGGATGTCCA	AAACCGCTTT	AAAATCGCCG	AAGCGTCTTT	TTATGCCGAG
	301	GAGTACGTCC	GTTTTCTGGA	ACGTTCGGAA	ACGGTTTCCG	AAAGCGCACT
	351	GCGTCAGTTT	TATGAGCGGC	AAATCCGCAT	GATCAAATTG	CAGCAGGTCA
	401	GCTTCGCAAC	CGAAGAGGAG	GCGCGTCAGG	CGCAGCAGCT	CCTGCTCAAA
	451	GGGCTGTCTT	TTGAAGGGCT	GATGAAGCGT	TATCCGAACG	ACGAGCAGGC
30	501	GTTTCGACGGT	TTCATTATGG	CGCAGCAGCT	TCCCGAGCCG	CTGGCTTcgc
	551	agtttgCCGG	TATGAACCGT	GGCGACGTTA	CCCGCAATCC	GGTCAAATTG
	601	GGCGAACGCT	ATTACCTGTT	CAAACTCGGC	GCGGTCGGGA	AAAACCCCGA
	651	CGCGCAGCCT	TTCGAGTTGG	TCAGAAACCA	GTTGGAACAA	GGTTTGAGGC
35	701	AGGAAAAAGC	CCGCTTGAAA	ATCGATGCCC	TTTTGGAaga	Aaacggtgtc
	751	AaacCGTAA				

This encodes a protein having amino acid sequence [<SEQ ID 302>] (SEQ ID NO: 302):

	1	<u>MKQKKTAAAV</u>	<u>IAAMLAGFAA</u>	<u>AKAPEIDPAL</u>	VDTLVAQIMQ	QADRHAEQSQ
	51	RPDQGAIRND	AVRRLQTLV	LKNRALKEGL	DKDKDVQNR	KIAEASFYAE
40	101	EYVRFLERSE	TVSESALRQF	YERQIRMIKL	QVVSFATEEE	ARQAQQLLLK
	151	GLSFEGLMKR	YPNDEQAFDG	FIMAQQLPEP	LASQFAGMNR	GDVTRNPVKL
	201	GERYYLFLKG	AVGKNPDAQP	FELVRNQLEQ	GLRQEKARLK	IDALLEENG
	251	KP*				

ORF76ng (SEQ ID NO: 302) and ORF76-1 (SEQ ID NO: 298) show 96.0% identity in 252 aa overlap

		10	20	30	40	50	60
	orf76-1.pep	MKQKKTAAAVIAAML	AGFAAAKAPEIDPAL	VDTLVAQIMQQA	DRHAEQSQKPDG	QAIRND	
5	orf76ng	MKQKKTAAAVIAAML	AGFAAAKAPEIDPAL	VDTLVAQIMQQA	DRHAEQSQRPDG	QAIRND	
		10	20	30	40	50	60
	orf76-1.pep	AVRRLQ	TLEVLKNRALKEGLD	KDKDVQNRFKIAEAS	FYAE EYVRFLERSETV	SEDELHKF	
10	orf76ng	AVRRLQ	TLEVLKNRALKEGLD	KDKDVQNRFKIAEAS	FYAE EYVRFLERSETV	SESALRQF	
		70	80	90	100	110	120
	orf76-1.pep	YEQQIRMIK	LQQVSFATEEEARQA	QQLLLKGLSF	EGLMKRYPNDEQA	FDGFIMAQQLPEP	
15	orf76ng	YERQIRMIK	LQQVSFATEEEARQA	QQLLLKGLSF	EGLMKRYPNDEQA	FDGFIMAQQLPEP	
		130	140	150	160	170	180
	orf76-1.pep	LASQFAAMNR	GDVTRDPVKLG	ERYFLFKLSEVG	KNPDAQPFELVR	NQLEQGLRQE	KARLK
20	orf76ng	LASQFAGMNR	GDVTRNPNVKLG	ERYFLFKLSEVG	KNPDAQPFELVR	NQLEQGLRQE	KARLK
		190	200	210	220	230	240
	orf76-1.pep	IDALLEENG	VKPKX				
25	orf76ng	IDALLEENG	VKPKX				
		250					

Furthermore, ORF76ng (SEQ ID NO: 302) shows significant homology to a *B.subtilis* export protein precursor (SEQ ID NO: 1132):

30	sp P24327 PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR)gi 98227 pir S15269 33K lipoprotein - Bacillus subtilis)gi 39782 (X57271) 33kDa lipoprotein [Bacillus subtilis]
)gi 2226124 gnl PID e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
)gi 2633331 gnl PID e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
35	Length = 292 Score = 50.4 bits (118), Expect = 1e-05 Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)
	Query: 70 VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAE EYVRFLERSETVSE 114 VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
	Sbjct: 53 VLTQLVQEKVLDKKYKVSDKEIDNKLKEYKTQLGDQYTALEKQYGKDYLKEQVKYELLTQ 112
40	Query: 115 SA-----LRQFYERQIRMIK LQQVSFATEEEARQAQQLLLKGLSF EGLMKRYPN 163 A +++++E I+ + A ++ A + ++ L KG FE L K Y
	Sbjct: 113 KAAKDNIVTDADIKEYWEGLK GKIRASHILVADKKTAEEVEKKLKG EKGFEDLAKEYST 172
	Query: 164 DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLG ERYFLFKLSEVGKNPDA 218 D A G F Q+ E + + G+V+ DPVK Y++ K +E D
45	Sbjct: 173 DSSASKGGDLGWFAKEGQMDETFSKAAFLKTGEVS-DPVKTQYGYHI IKKTEERGKYDD 231
	Query: 219 QPFELVRNQLEQGLRQEKA 237 EL LEQ L A
	Sbjct: 232 MKKELKSEVLEQKLNDNAA 250

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (SEQ ID NO: 298) (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 (SEQ ID NO: 298) is a surface-exposed protein, and that it is a useful immunogen.

10 Example 36

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 303] (SEQ ID NO: 303):

```

1  ATGAAAAAAT CTTTCCTTAC GCTTGTTCTG TATTCGTCTT TACTTACCGC
51  CAGCGAAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG
15  101  GCAAAAATTG CCGAAACGTT TGCCTGACA TTTGTGATTG CTGCGCTGTA
151  TCTGTTTGCG CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
201  CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
251  ATGACG. ....//
20  1201  .... CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
1251  ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
1301  ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT
1351  GTTGTTACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
1401  CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCAACGTT CCTGATTAC
25  1451  ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTCGCGAAG GCTCGGTAAC
1501  GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA
1551  AGGCGGAATA TGTTTATCCG CAATGA

```

This corresponds to the amino acid sequence [SEQ ID 304; ORF81] (SEQ ID NO: 304; ORF81):

```

1  MKKSFLTLVL YSLLLTASEI AYPLELGIET LPAAKIAETF ALTFVIAALY
51  LFARNKVTRL LIAVFFAFSI IANNVHYADY QSWMT.....//
35  401  ...QTVEQL QKTPDGNWLF AYTSDHGQYV RQDIYNQGTQ QPDSYLVPLV
451  LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGVDMP VSGCREGSVT
501  GNLTGDAGS LNIRDGKA EY VYPQ*

```

Further work revealed the complete nucleotide sequence [SEQ ID 305] (SEQ ID NO: 305):

```

1  ATGAAAAAAT CTTTCCTTAC GCTTGTTCTG TATTCGTCTT TACTTACCGC

```


51 CAGCGAAAATT GCCTATCGCT TTGTATTTGG GATTGAAACC TTACCGGCGG
 101 CAAAAATTGC GGAAACGTTT GCGCTGACAT TTGTGATTGC TGCGCTGTAT
 151 CTGTTTGCGC GTTATAAGGT GACGCGTTTG TTGATTGCGG TGTTTTTTGC
 201 GTTCAGCATT ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
 5 251 TGACGGGCAT CAATTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
 301 AGCGCGGGTG CGTCGATGTT GGATAAGTTG TGGCTGCCTG TGTGTGGGG
 351 CGTGTGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
 401 CGCATTTTTT TGCCGATATA CTGTTTGCTT TCCTAATGCT GATGATTTTC
 451 GTGCGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTTCG CCAAACCGAC
 10 501 ATACAGCCGC ATCAAAGCCA ATTATTTTCAG CTTGCGTTAT TTGTGCGGAC
 551 GCGTGTTGCC GTATCAGTTG TTTGATTTAA GCAGGATTCC CGCCTTTAAG
 601 CAGCTGCTC CAAGCAAAAT CGGGCAGGGC AGTGTTCAAA ATATCGTCTT
 651 GATTATGGGC GAAAGCGAAA GCGCGCGCA TTTGAAGCTG TTGGGCTACG
 701 GACGCGAAAC TTCGCCGTTT TTAACCCGGC TGTCGCAAGC CGATTTTAAG
 15 751 CCGATTGTGA AACAAAGTTA TTCCGCAGGC TTTATGACTG CAGTGTCCCT
 801 GCCCAGTTTT TTCAATGCGA TACCGCACGC CAACGGCTTG GAACAAATCA
 851 GCGGCGGCGA TACCAATATG TTCCGCCTCG CCAAAGAGCA GGGCTATGAA
 901 ACGTATTTTT ACAGCGCGCA GGCGGAAAAC GAGATGGCGA TTTTGAACTT
 951 AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTGGCT
 20 1001 ACGGCAACGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
 1051 AAAATCAATT TGCAGCAGGG CAAGCATTTT ATCGTGTTGC ACCAACGCGG
 1101 TTCGCACGCC CCATACGGCG CATTGTTGCA GCCTCAAGAT AAAGTATTCT
 1151 GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
 1201 CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
 25 1251 CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTT CGCCAAGATA
 1301 TCTACAATCA AGGCACGGTG CAGCCGACA GCTATCTCGT GCCGCTAGTG
 1351 TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
 1401 GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTACAA
 1451 CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGGAAGG CTCGGTAACG
 30 1501 GGCAACCTGA TTACGGGTGA TGCAGGCAGC TTGAACATTC GCGACGGCAA
 1551 GGCGGAATAT GTTTATCCGC AATGA

This corresponds to the amino acid sequence [<SEQ ID 306; ORF81-1>] (SEQ ID NO: 306; ORF81-1):

35 1 MKKSFLTLVL YSLLTASEI AYRFVFGIET LPAAKIAETF ALTFVIAALY
 51 LFARYKVTRL LIAVFFAFSI IANNVHYAVY QSWMTGINYW LMLKEVTEVG
 101 SAGASMLDKL WLPVLWGVLE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
 151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSRIPAFK
 201 QPAPSKIQQG SVQNIIVLIMG ESESAHLKL FGYGRETSFP LTRLSQADFK
 40 251 PIVKQSYSAG FMTAVSLPSF FNAIPHANGL EQISGGDTNM FRLAKEQGYE
 301 TYFYSAQEN EMAILNLIGK KWIDHLIPT QLGYNGDNM PDEKLLPLFD
 351 KINLQQGKHF IVLHQRGSHA PYGALLQPQD KVFGEADIVD KYDNTIHKTD
 401 QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
 451 LYSYPAKAVQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
 45 501 GNLITGDAGS LNIRDGAKEY VYPQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 (SEQ ID NO: 304) shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) (SEQ ID NO: 308) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf81.pep	MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALTFVIAALYLFARNKVTRL					
	orf81a	MKKSFLVLFYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL					
5		10	20	30	40	50	60
	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT					
	orf81a	LIAVFFAFSIIANNVHYAVYQSWITGINYWLMLKEITEVGGAGASMLDKLWLPALWGVLE					
10		70	80	90	100	110	120
		//					
	orf81.pep	QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD					
15	orf81a	IPHANGLEQISGGDIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD					
		280	290	300	310	320	330
	orf81.pep	IYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG					
20	orf81a	IYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG					
		340	350	360	370	380	390
	orf81.pep	CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
25	orf81a	CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
		400	410	420			

The complete length ORF81a nucleotide sequence [<SEQ ID 307>] (SEQ ID NO: 307) is:

	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTCTG	TATTCGTCCC	TACTTACTGC
30	51	CAGCGAAATT	GCTTATCGCT	TTGTATTCGG	AATTGAAACC	TTACCGGCTG
	101	CAAAAATGGC	AGAAACGTTT	GCGCTGACAT	TTGTGATTGC	TGCGCTGTAT
	151	CTGTTTGCGC	GTTATAAGGC	AACGCGTTTG	TTGATTGCGG	TGTTTTTCGC
	201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
35	251	TAACGGGCAT	TAATTATTGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
	301	GGCGCAGGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CGTTGTGGGG
	351	CGTGTTGGAA	GTCATGTTGT	TTTGACGCTT	TGCCAAGTTC	CGCCGTAAGA
	401	CGCATTTTTC	TGCCGATATA	CTGTTTGCCT	TCCTAATGCT	GATGATTTTC
	451	GTGCGTTCGT	TCGACACGAA	ACAAGAACAC	GGTATTTTCG	CCAAACCGAC
40	501	ATACAGCCGC	ATCAAAGCCA	ATTATTTCAG	CTTCGGTTAT	TTTGTCCGAC
	551	GCGTGTTGCC	GTATCAGTTG	TTTGATTAA	GCAAGATTCC	TGTGTTCAAA
	601	CAGCCTGCTC	CAAGCAGAA	CGGGCAAGGC	AGTATTCAAA	ATATCGTCCT
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGCTACG
	701	GGCGCGAAAC	TTGCGCGTTT	TTGACCCAGC	TTTCGCAAGC	CGATTTTAA
45	751	CCGATTGTGA	AACAAAGTTA	TTCCGCAGGC	TTTATGACGG	CAGTATCCCT
	801	GCCCAGTTTC	TTTAACGTCA	TACCGCATGC	CAACGGCTTG	GAACAAATCA
	851	GCGGCGGCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAACCGGAC
	901	CAAATGATTC	AAACCGTATT	CGAGCAGCTG	CAAAGCAGC	CTGACGGCAA
	951	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTT	CGCCAAGATA
50	1001	TCTACAATCA	AGGACGGTG	CAGCCGACA	GCTATCTCGT	GCCGCTGGTG
	1051	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCCAACC	AGGCTTTTGC
	1101	GCCTTGCGAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTCA
	1151	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCGCGAAGG	CTCGGTAACG
	1201	GGCAACCTGA	TTACGGGTGA	TGCAGGCAGC	TTGAACATTC	GCGACGGCAA
55	1251	GGCGGAATAT	GTTTATCCGC	AATGA		

This encodes a protein having amino acid sequence [<SEQ ID 308>] (SEQ ID NO: 308):

```

1  MKKSLFVLFL YSSLLTASEI AYRFVFGIET LPAAKMAETF ALTFVIAALY
51  LFARYKATRL LIAVFFAFSI IANNVHYAVY QSWITGINYW LMLKEITEVG
101 GAGASMLDKL WLPALWGVLE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
5  151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSKIPVFK
201 QPAPSRIGQG SIQNIVLIMG ESESAHLKL FGYGRETSFP LTQLSQADFK
251 PIVKQSYSAG FMTAVSLPSF FNVIPHANGL EQISGGDIVD KYDNTIHKTD
301 QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
10 351 LYSPPKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
401 GNLTG DAGS LNIRDGKAEY VYPQ*

```

ORF81a (SEQ ID NO: 308) and ORF81-1 (SEQ ID NO: 306) show 77.9% identity in 524 aa overlap:

```

15  orf81a.pep      10      20      30      40      50      60
    orf81-1      MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL
    |||:::| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf81-1      MKKSFLTLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL
    10      20      30      40      50      60

20  orf81a.pep      70      80      90      100     110     120
    orf81-1      LIAVFFAFSI IANNVHYAVYQSWITGINYWMLMLKEITEVGGAGASMLDKLWLPALWGVLE
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf81-1      LIAVFFAFSI IANNVHYAVYQSWMTGINYWMLMLKEVTEVGSAGASMLDKLWLPVLWGVLE
    70      80      90      100     110     120

25  orf81a.pep      130     140     150     160     170     180
    orf81-1      VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSR IKANYFSFGY
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf81-1      VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSR IKANYFSFGY
    130     140     150     160     170     180

30  orf81a.pep      190     200     210     220     230     240
    orf81-1      FVGRVLPYQLFDLSKIPVFKQPAPSRIGQGS IQNIVLIMGESESAHLKLFGYGRETSFP
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf81-1      FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIVLIMGESESAHLKLFGYGRETSFP
    190     200     210     220     230     240

35  orf81a.pep      250     260     270     280
    orf81-1      LTQLSQADFKPIVKQSYSAGFMTAVSLPSFFNVIPHANGLEQISGGD-----
    ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf81-1      LTRL SQADFKPIVKQSYSAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFRLAKEQGYE
    250     260     270     280     290     300

40  orf81a.pep      -----
    orf81-1      TYFYSAQAENEMAILNLI GK KWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF
    310     320     330     340     350     360

45  orf81a.pep      ----- 290 300 310 320
    orf81-1      ----- IVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    orf81-1      IVLHQRGSHAPYGALLQPQDKVFEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF
    370     380     390     400     410     420

```

		330	340	350	360	370	380
5	orf81a.pep	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
	orf81-1	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
	orf81a.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
10	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
		490	500	510	520		

Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 (SEQ ID NO: 304) and a predicted ORF (ORF81.ng) (SEQ ID NO: 310) from *N. gonorrhoeae* of the N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

15	orf81.pep	MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALTFFVIAALYLFARNKVTRL	60
	orf81ng	MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFMIAALYLFARYKASRL	60
	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT	85
20	orf81ng	LIAVFFAFSMIANNVHYAVYQSWMTGINYWLMLKEVTEVGSAGASMLDKLWLPALWGVAE	120
	orf81.pep	QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD	433
	orf81ng	ALLQPQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD	433
25	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81ng	IYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81.pep	CREGSVTGNLITGDAGSLNIRDGKAEYVYPQ	524
30	orf81ng	CREGSVTGNLITGDAGSLNIRNGKAEYVYPQ	524

The complete length ORF81ng nucleotide sequence [<SEQ ID 309>] (SEQ ID NO: 309) is:

	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCTG	TATTCATCCC	TACTTACCGC
35	51	CAGCGAAATC	GCCTATCGCT	TTGTATTCGG	AATTGAAACC	TTACCGGCTG
	101	CAAAAATGGC	GGAAACGTTT	GCGCTGACAT	TTATGATTGC	TGCGCTGTAT
	151	CTGTTTGCGC	GTTATAAGGC	TTGCGGCTG	CTGATTGCGG	TGTTTTTCGC
	201	GTTCAGCATG	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
	251	TGACGGGTAT	TAACATTGG	CTGATGCTGA	AAGAGGTTAC	CGAAGTCGGC
	301	AGCGCGGGCG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CTTTGTGGGG
40	351	CGTGGCGGAA	GTCATGTTGT	TTTGCAGCCT	TGCCAAGTTC	CGCCGTAAGA
	401	CGCATTTTTC	TGCCGATATA	CTGTTTGCCT	TCCTAATGCT	GATGATTTTC
	451	GTGCGTTTCG	TCGACACGAA	ACAAGAGCAC	GGTATTTTCG	CCAAACCGAC
	501	ATACAGCCGC	ATCAAAGCCA	ATTATTTTCA	CTTCGGTTAT	TTTGTGCGGC
	551	GCGTGTGTC	GTATCAGTTG	TTTGATTAA	GCAAGATCCC	TGTGTTCAAA
45	601	CAGCCTGCTC	CAAGCAAAAT	CGGGCAAGGC	AGTATTCAAA	ATATCGTCC
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGTTACG
	701	GGCGCGAAAC	TTGCGCGTTT	TTAACCCGGC	TGTCGCAAGC	CGATTTTAAG

5	751	CCGATTGTGA	AACAAAGTTA	TTCCGCAGGC	TTTATGACGG	CAGTATCCCT
	801	GCCCAGTTTC	TTTAACGTCA	TACCGCACGC	CAACGCCTTG	GAACAAATCA
	851	GCGGCGGCGA	TACCAATATG	TTCCGCCTCG	CCAAAGAGCA	GGGCTATGAA
	901	ACGTATTTTT	ACAGTGCCCA	GGCTGAAAA	CAAAATGGCA	TTTTGAACTT
	951	AATCGGTAAG	AAATGGATAG	ACCATCTGAT	TCAAGCCAGC	CAACTTGGCT
10	1001	ACGGCAACGG	GCACAATATG	CCCGATGAGA	AGCTGCTGCC	GTGTGTCGAC
	1051	AAAATCAATT	TGCAGCAGGG	CAGGCATTTT	ATCGTGTTC	ACCAACGCGG
	1101	TTCGCACGCC	CCATACGCGC	CATTGTTGCA	GCCTCAAGAT	AAAGTATTTC
	1151	GCGAAGCCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCGAC
	1201	CAAAATGATT	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
15	1251	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTG	CGCCAAGATA
	1301	TCTACAATCA	AGGCACGGTG	CAGCCCGACA	GCTATATTGT	GCCTCTGGTT
	1351	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCCAACC	AGGCTTTTGC
	1401	GCCTTGCGAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTACAA
	1451	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCCGGAAGG	CTCGGTAACA
	1501	GGCAACCTGA	TTACGGGCGA	TGCAGGCAGC	TTGAACATTG	GCAACGGCAA
	1551	GGCGGAATAT	GTTTATCCGC	AATAA		

This encodes a protein having amino acid sequence [<SEQ ID 310>] (SEQ ID NO: 310):

20	1	MKKSFLVFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFMIAALY
	51	<u>LFARYKASRL</u>	<u>LI</u> AVFFAFSM	IANNVHYAVY	QSWMTGINYW	LMLKEVTEVG
	101	SAGASMLDKL	WLPALWGVAE	VMLFCSLAKF	RRKTHPSADI	LFAFLMLMIF
	151	<u>VRSFDTKQEH</u>	GISPKPTYSR	IKANYFSFGY	FVGRVLPYQL	FDLSKIPVFK
	201	QPAPSKIGQG	SIQNIIVLIMG	ESESAAHLKL	FGYGRETSPP	LTRLSQADFK
25	251	PIVKQSYSAG	FMTAVSLPSF	FNVIPHANGL	EQISGGDTNM	FRLAKEQGYE
	301	TYFYSAQAE	QMAILNLIGK	KWIDHLIQPT	QLGYNGDNM	PDEKLLPLFD
	351	KINLQQGRHF	IVLHQGRSHA	PYGALLQPQD	KVFGEADIV	KYDNTIHKTD
	401	QMIQTVEEQ	KQKPDGNWLF	AYTSDHGQYV	RQDIYNQGT	QPDSDYIVPLV
	451	LYSPDKAVQQ	AAANAFAPCE	IAFHQQLSFT	LIHTLGYDMP	VSGCREGSVT
30	501	GNLITGDAGS	LNIRNGKAEY	VYPQ*		

ORF81ng (SEQ ID NO: 310) and ORF81-1 (SEQ ID NO: 306) show 96.4% identity in 524 aa overlap:

			10	20	30	40	50	60
35	orf81ng-1.pep	MKKS	LFVLFLYSSLLT	ASEIAYRFVFGIETL	PAAKMAET	FALTFMIAALYLF	FARYKASRL	
	orf81-1	MKKS	SFLTFLVLYSSLLT	ASEIAYRFVFGIETL	PAAKIAET	FALTFVIAALYLF	FARYKVTRL	
			10	20	30	40	50	60
			70	80	90	100	110	120
40	orf81ng-1.pep	LI	AVFFAFSMIANNVHYAVYQSWMTG	INYLMLKEVTEVGSAGASMLDKLWLP	ALWGVAE			
	orf81-1	LI	AVFFAFSIIANNVHYAVYQSWMTG	INYLMLKEVTEVGSAGASMLDKLWLP	VLWGVLE			
			70	80	90	100	110	120
			130	140	150	160	170	180
45	orf81ng-1.pep	VMLF	CSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTY	SRIKANYFSFGY				
	orf81-1	VMLF	CSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTY	SRIKANYFSFGY				
			130	140	150	160	170	180
			190	200	210	220	230	240
50	orf81ng-1.pep	FVGR	VLPLYQLFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESA	AHLKLF	FGYGRETS	PF		
	orf81-1	FVGR	VLPLYQLFDLSRIPAFKQPAPSKIGQGSVQNIVLIMGESESA	AHLKLF	FGYGRETS	PF		

-270-

		190	200	210	220	230	240
		250	260	270	280	290	300
	orf81ng-1.pep	LTRL SQADFKPIVKQSY SAGFMTAVSLPSFFNVIPHANGLEQISGGDTNMFR LAKEQGYE					
5	orf81-1	LTRL SQADFKPIVKQSY SAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFR LAKEQGYE					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf81ng-1.pep	TYFYSAQAENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGRHF					
10	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf81ng-1.pep	IVLHQRGSHAPYGALLQPQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF					
15	orf81-1	IVLHQRGSHAPYGALLQPQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF					
		370	380	390	400	410	420
		430	440	450	460	470	480
	orf81ng-1.pep	AYTSDHGQYVRQDIYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQQLSTF					
20	orf81-1	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQQLSTF					
		430	440	450	460	470	480
		490	500	510	520		
	orf81ng-1.pep	LIHTLGYDMPVSGCREG SVTGNLITGDAGSLNIRNGKAEYVYPQX					
25	orf81-1	LIHTLGYDMPVSGCREG SVTGNLITGDAGSLNIRDGKAEYVYPQX					
		490	500	510	520		

Furthermore, ORF81ng (SEQ ID NO: 310) shows significant homology to an *E.coli* OMP (SEQ ID NO: 1133):

30	gi 1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)	
35	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIAVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHMOVQRLLFFVLTLVVKRISSLPLRLLVAAPFVL-LTAADMSISLY- 86	
	Query: 82 SWMT-----GINYWMLMLKEVTEVGSAGASMLDKLWLPALWGVAEVMFLFCSLAKFRRT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV---AKMLG-MYSPYLCAFAFLSLLFLAVI IKYDV 141	
40	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISP KPTYSRIKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKKVTGILLLIVISGSLFSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201	
45	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQSGIQNIVLIMGESESAHLKLFYGRETSPFL 241 +Q L + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVPYFQL---SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257	
	Query: 242 TRLSQADFKPIVKQSY SAGFMTAVSLP---SFFNVIPHANGLEQISGGDTNMFR LAKEQG 298	

+Q + Q+ S TA+S+P + +V+ H I N+ +A + G
 Sbjct: 258 E--AQRKQIKLFNQAISGAPYTALSVPLSLTADSVLSH-----DIHNPDPNIINMANQAG 310
 Query: 299 YETYFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355
 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q
 5 Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLPHLSQALQQ 359
 Query: 356 --QGRHFIVLHQRGSHAPYGALLQPQDKVFGADIVDK-YDNTIHKTDQMIQTVFEQLQK 412
 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+
 Sbjct: 360 NTQQKKLIVLHLNGSHEPACSAYPQSSAVFQPQDDQDACYDNSIHYTDSLQGQVFELLK- 418
 Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQG--TVQPDSYIVPL-VLYSP 454
 D Y +DHG ++++Y G +Y VP+ + YSP
 10 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVPFIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined)
 and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is
 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 15 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 37

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 311>] (SEQ ID
 NO: 311):

20 1 ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGC GACACTGAC
 51 CGGCATACTC GCCCaCGGCG GCGGCAAACG CTTTGCCGTC GAACAAGAAC
 101 TCGTCGCCGC ATCGTCCCGC GCCGCCGTCA AAGAAATGGA TTTGTCCGCC
 151 yTAAAAGGAC GCAAAGCCGC CyTTTACGTC TCGTTATGG GCGACCAAGG
 201 TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
 25 251 GCGGCTACCA CAACAACCCC GAAAGTGCCA CCCAATACAG CTACCCCGCC
 301 TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
 351 TTCCACATCG CTTTGAACG CCCCCGCCGC CGyCyTGACG AAAACAGCG
 401 GACGCAAAGG CGAACGcTCC GCCGGAAGT CCGTCAACGG CACGGGCGAC
 451 TACCGCAACG AAACCCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
 501 CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
 30 551 CCGrATACGC CGACACCGAC GTATTCGTAA CCGTCGACGT A...

This corresponds to the amino acid sequence [<SEQ ID 312; ORF83>] (SEQ ID NO: 312;
 ORF83):

35 1 ..TLLFIPLVL TXCGTLTGIL AHGGGKRFAV EQELVAASSR AAVKEMDLA
 51 LKGRKAAXYV SVMGDQGSN ISGGYRYSIDA LIRGGYHNNP ESATQYSYPA
 101 YDTTATTKSD ALSSVTTSTS LLNAPAAXLT KNSGRKGERS AGLSVNGTGD
 151 YRNETLLANP RDVSFLTNI QTVFYLRGIE VVPPXYADTD VFVTVDV..

Further work revealed the complete nucleotide sequence [<SEQ ID 313>] (SEQ ID NO: 313):

40 1 ATGAAAACCC TGCTCCTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC

1	MKTLLLLLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
51	SALKGRKAAL	YVSMGDQGS	GNISGGRYSI	DALIRGGYHN	NPESATQYSY
101	PAYDTTATTK	SDALSSVTTT	TSLLNAPAAA	LTKNNGRKGGE	RSAGLSVNGT
151	GDYRNETLLA	NPRDVSFLT	LIQTVFYLRG	IEVVPPEYAD	TDVFTVDFV
201	QTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLITPK	TAAYESQYQE
251	GVALWTGPYK	VSKTVKASDR	LMVDFSDITP	YGDTTAQNRP	DFKQNGKKP
301	DVGNEVIRRR	KGG*			

35	orf83.pep	<u>TLLLF</u> IPLVLTXCGTLTGILAHGGGKRF	AVEQELVAASSRAAVKEMDLSALKGRKAAX
	orf83a	<u>MKTLLXL</u> IPLVLTA	CGTLTGIPAHGGGKRF
			AVEQELVAASSRAAVKEMDLSALKGRKAAL
		10	20
		30	40
		50	60
40	orf83.pep	YVSMGDQGS	GNISGG
	orf83a	YVSMGDQGS	GNISGG
		70	80
		90	100
		110	120
45	orf83.pep	TSLLNAPAA	XLTKNSGRK
	orf83a	TSLLNAPAA	AALTKNSGRK
		130	140
		150	160
		170	180

180 190
orf83.pep IEVVPPXYADTDVFVTVDV
||||| |||||||||
orf83a IEVVPEYADTDVFVTVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDRSRKLLIAPK
5 190 200 210 220 230 240

The complete length ORF83a nucleotide sequence [<SEQ ID 315>] (SEQ ID NO: 315) is:

1 ATGAAAACCC TGCTCNTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
51 ACTGACCGGC ATACCCGCC ACGGCGGCGG CAAACGCTTT GCCGTCGAAC
10 101 AAGAACTCGT CGCCGCATCG TCCCGCGCCG CCGTCAAAGA AATGGACTTG
151 TCCGCCCTGA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
201 CCAAGGTTCT GGCAACATAA GCGGCGGACG CTACTCTATC GACGCACTGA
251 TACGCGGCGG CTACCACAAC AACCCCGAAA GTGCCACCCA ATACAGCTAC
301 CCCGCCCTACG AACTACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT
15 351 AACCATTTC ACATCGCTTT TGAACGCCCC CGCCGCGGCC CTGACGAAAA
401 ACAGCGGACG CAAAGGCGAA CGCTCCGCCG GACTGTCCGT CAACGGCACG
451 GCGGACTACC GCAACGAAAC CCTGCTCGCC AACCCCGCG ACGTTTCCTT
501 CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
551 TACCGCCCGA ATACGCCGAC ACCGACGTAT TCGTAACCGT CGACGTATTTC
20 601 GGCACCGTCC GCAGCCGCAC CGAAGTCAC CTCTACAACG CCGAAACCTT
651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCCGGA
701 AACTGCTGAT TGCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA
751 CAATACGCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC
801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCCCC TACGGCGACA
25 851 CAACCGCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCC
901 GATGTCGGCA ACGAAGTCAT CCGCCGCCG AAAGGAGGAT AA

This encodes a protein having amino acid sequence [<SEQ ID 316>] (SEQ ID NO: 316):

1 MKTLLXLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
30 51 SALKGRKAAL YVSMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY
101 PAYDTTATTK SDALSSVTTS TSLLNAPAAA LTKNSGRKGE RSAGLSVNGT
151 GDYRNETLLA NPRDVSFLTN LIQTVFYLRG IEVVPPEYAD TDVFVTVDVF
201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE
251 QYALWMPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNGKKP
35 301 DVGNEVIRRR KGG*

ORF83a (SEQ ID NO: 316) and ORF83-1 (SEQ ID NO: 314) show 98.4% identity in 313 aa overlap:

10 20 30 40 50 60
40 orf83a.pep MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL
||||| |||||||||
orf83-1 MKTLLLLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL
10 20 30 40 50 60
70 80 90 100 110 120
45 orf83a.pep YVSMGDQGS GNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS
||||| |||||||||
orf83-1 YVSMGDQGS GNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS
70 80 90 100 110 120
130 140 150 160 170 180
50 orf83a.pep TSLLNAPAAALTKNNGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNLIQTVFYLRG
||||| |||||||||

-274-

5	orf83-1	TSLLNAPAAALTKN	SGRK	GERS	SAGLSV	NGTGDY	RNETLL	ANPRDV	SFLT	NLIQT	VFYLRG
		130	140	150	160	170	180				
10	orf83a.pep	IEVVPPEYADTDV	FVTVDV	FGTVRS	TELHLY	NAETLK	AQTKLE	YFAVDR	DSRKLL	IAPK	
	orf83-1	IEVVPPEYADTDV	FVTVDV	FGTVRS	TELHLY	NAETLK	AQTKLE	YFAVDR	DSRKLL	ITPK	
15	orf83a.pep	TAAYESQYQEYAL	WMGPYS	VGKTVK	ASDRLM	VDFS	DITPYG	DTTAQN	RPDFKQ	NNGKKP	
	orf83-1	TAAYESQYQEYAL	WTGPYK	VSKTVK	ASDRLM	VDFS	DITPYG	DTTAQN	RPDFKQ	NNGKKP	
	orf83a.pep	DVGNEVIRRRKGGX									
	orf83-1	DVGNEVIRRRKGGX									

Homology with a predicted ORF from *N. gonorrhoeae*

ORF83 (SEQ ID NO: 312) shows 94.9% identity over a 197aa overlap with a predicted ORF
 20 (ORF83.ng) (SEQ ID NO: 318) from *N. gonorrhoeae*:

25	orf83.pep	TLLLFIPLVLTXC	GTLTGIL	AHGGGK	RFAVEQ	ELVAASS	RAAVK	EMDLS	ALKGR	KAAX	58
	orf83ng	MKTLLLLIPLVLT	ACGTLTG	IPAHGGG	KRFAVEQ	ELVAASS	RAAVK	EMDLS	ALKGR	KAAL	60
30	orf83.pep	YVSVMGDQGS	GNISGG	RYSIDA	LIRGGY	HNNPES	ATQYSY	PAYD	TTATT	KSDALS	SVTTS 118
	orf83ng	YVSVMGDQGS	GNISGG	RYSIDA	LIRGGY	HNNPDS	ATRYSY	PAYD	TTATT	KSDALS	SGVTTS 120
	orf83.pep	TSLLNAPAA	XLTKNS	GRKGER	SAGLSV	NGTGDY	RNETLL	ANPRDV	SFLT	NLIQT	VFYLRG 178
	orf83ng	TSLLNAPAA	ALTKNNG	RKGER	SAGLSV	NGTGDY	RNETLL	ANPRDV	SFLT	NLIQT	VFYLRG 180
	orf83.pep	IEVVPXYADTD	VFVTVDV								197
	orf83ng	IEVVPPEYADTD	VFVTVDV	FGTVRS	TELHLY	NAETLK	AQTKLE	YFAVDR	DSRKLL	IAPK	240

The complete length ORF83ng nucleotide sequence [<SEQ ID 317>] (SEQ ID NO: 317) is:

35	1	ATGAAAACCC	TGCTCCTCCT	CATCCCCCTC	GTACTCACCG	CCTGCGGCAC	
	51	ACTGACCGGC	ATACCCGCCC	ACGGCGGCGG	CAAACGCTTT	GCCGTCGAAC	
40	101	AGGAACTCGT	CGCCGCATCG	TCCCGCGCCG	CCGTCAAAGA	AATGGACTTG	
	151	TCCGCCCTGA	AAGGACGCAA	AGCCGCCCTT	TACGTCTCCG	TTATGGGCGA	
	201	CCAAGGTTTC	GGCAACATAA	GCGGCGGACG	CTACTCCATC	GACGCACTGA	
	251	TACGCGGCGG	CTACCACAAC	AACCCCGACA	GCGCCACCCG	ATACAGCTAC	
45	301	CCCGCCTATG	ACACTACCGC	CACCACCAAA	TCCGACGCGC	TCTCCGGCGT	
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCCGCC	CTGACGAAAA	
	401	ACAACGGACG	CAAAGGCGAA	CGTCCGCCG	GACTGTCCGT	CAACGGCACG	
	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCGCG	ACGTTTCCTT	
	501	CCTGACCAAC	CTCATCCAAA	CCGTCTTCTA	CCTGCGCGGC	ATCGAAGTCG	

-275-

551 TACCGCCCGA ATACGCCGAC ACCGACGTAT TCGTAACCGT CGACGTATTC
601 GGCACCGTCC GCAGCCGTAC CGAACTGCAC CTCTACAACG CCGAAACCCCT
651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTCGACCGC GACAGCCGGA
701 AACTGCTGAT TGCCCCATAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA
5 751 CAATACGCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC
801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCCCC TACGGCGACA
851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCCC
901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

10 This encodes a protein having amino acid sequence [<SEQ ID 318>] (SEQ ID NO: 318):

1 MKTLLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
51 SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPDSATRYSY
101 PAYDTTATTK SDALSGVTTs TSLNAPAAA LTKNNGRKGE RSAGLSVNGT
15 151 GDYRNETLLA NPDVSVFLTN LIQTVFYLRG IEVVPPEYAD TDVFTVDFV
201 GTVRSRTELH LYNATLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE
251 QYALWMPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRPFDFKQNGKNP
301 DVGNEVIRRR KGG*

20 ORF83ng (SEQ ID NO: 318) and ORF83-1 (SEQ ID NO: 314) show 97.1% identity in 313 aa overlap

		10	20	30	40	50	60
orf83-1.pep	MKTLLLLLIPLVLTACGTLTGIPAHGGGKRF	AVEQELVAASSRAAVKEMDLSALKGRKAAL					
orf83ng	MKTLLLLLIPLVLTACGTLTGIPAHGGGKRF	AVEQELVAASSRAAVKEMDLSALKGRKAAL					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf83-1.pep	YVSVMGDQGS	GNISGGRYSIDALIRGGYHN	NPESATQYSYPAYDTTATT	KSALSSVTTs			
orf83ng	YVSVMGDQGS	GNISGGRYSIDALIRGGYHN	NPDSATRYSY	PAYDTTATT	KSALSGVTTs		
		70	80	90	100	110	120
		130	140	150	160	170	180
orf83-1.pep	TSLNAPAAA	LTKNSGRK	GERSAGLSVNGT	GDYRNETLLAN	PRDVSFLTN	LIQTVFYLRG	
orf83ng	TSLNAPAAA	LTKNNGRK	GERSAGLSVNGT	GDYRNETLLAN	PRDVSFLTN	LIQTVFYLRG	
		130	140	150	160	170	180
		190	200	210	220	230	240
orf83-1.pep	IEVVPPEYADTDV	FVTVDVFGTVRSRTELH	LYNAETLKAQTKLEYFAVDR	DSRKLLITPK			
orf83ng	IEVVPPEYADTDV	FVTVDVFGTVRSRTELH	LYNAETLKAQTKLEYFAVDR	DSRKLLIAPK			
		190	200	210	220	230	240
		250	260	270	280	290	300
orf83-1.pep	TAAYESQYQE	QYALWTPYKVS	KTVKASDRLMVDFSDITPYGDTTAQNR	PDFKQNGKKP			
orf83ng	TAAYESQYQE	QYALWMPYS	VGKTVKASDRLMVDFSDITPYGDTTAQNR	PDFKQNGKNP			
		250	260	270	280	290	300
		310					
orf83-1.pep	DVGNEVIRRRKGGX						
orf83ng	DVGNEVIRRRKGGX						
		310					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from

5 *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 319>] (SEQ ID NO: 319):

```

10      1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
      51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
     101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
     151  CACACCTACA TAGAAACGGA CGCAAAAAG CTGCCGAAAT CGACAGATGA
     201  GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
     15  251  TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGGCACGC
     301  TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
     351  ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCCT AAGCTTCTAG
     401  ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACAT CGCTTCAAAC
     451  AAGATGGGTA TCGGTACGCT TTTAGAAATG AAAATATGCG CGGACGATCC
     20  501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
     551  AAGTTTATGA CTTGTATsrr TmmGCGGAAG TTCATACCGT AAATAAGGTC
     601  AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
     651  CGTGTTTGTG GGCCTGTCTT ATAAAATGTT GagCaGTTAC GGAAAAAAC
     701  aGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
     25  751  CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
     801  AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGrAAGC AAGCcgaTTT
     851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
     901  GAAGCGCGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
     951  gaAAGAAGTG ACGGaGTTGA TGTGcCaAgG aCTATGTaAA AAacGGCTTG
     30  1001  CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGCAGGAAG TTCAGCAAAG
     1051  CGGCAGCAA CATTCCGACA GGGCGcCAAG TTGCCACATT GGGCGGAAAA
     1101  CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAACCGG GGAAACCGTT
     1151  TGAAGGAATC GGaCGGGGGC GTGGTCGGAT CGGCAAACTG A

```

35 This corresponds to the amino acid sequence [<SEQ ID 320; ORF84>] (SEQ ID NO: 320; ORF84):

```

      1  MAEICLITGT PGSGKTLKMV SMMANDEMFK PDEKAIRRKV FTNIKGLKIP
     51  HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSVIV DEAQDVWPAR
     101  SAGSKIPENV QWLNTHRHQG IDIFVLTTQGP KLLDQNLRTL VRKHYHIASN
     40  151  KMGMRLLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVNVK
     201  KRSKWFTLP VIVLLIPVFV GLSYKMLSSY GKKQEEPAQ ESAATEQAV
     251  LPDKTEGEPV NNGNLTADMF VPTLSEKPKS KPIYNGVRQV RTFEYIAGCI
     301  EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGQEVQQS
     45  351  AQQHSDRAQV ATLGGKXPQN LMYDNWEERG KPFEGIGGGV VGSAN*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 321>] (SEQ ID NO: 321):

-277-

1 ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
 51 AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
 101 ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
 151 CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
 5 201 GCAGCTTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
 251 TCGGGTCTAT TGTCAATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
 351 ACATCAGGGC ATTGATATAT TTGTTTTGAC TCAAGGTCTT AAGCTTCTAG
 401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
 10 451 AAGATGGGTA TGCCTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
 501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
 551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
 601 AAGCGGTCAA AGTGGTTTTC CACTCTGCCA GTAATAGTAT TGCTGATTCC
 651 CGTGTGTTGTC GGCCTGTCTT ATAAAATGTT GAGCAGTTAC GGAAAAAAC
 15 701 AGGAAGAACC CGCAGCACA GAATCGGCGG CAACAGAACA GCAGGCAGTA
 751 CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
 801 AAGATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT
 851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
 901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGGCATT
 20 951 GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTGC
 1001 CGTTTAAACCC ATACAAAGAA GAAAGCCAAG GGCAGGAAGT TCAGCAAAGC
 1051 GCGCAGCAAC ATTCGGACAG GCGCAAGTT GCCACATTGG GCGGAAAACC
 1101 GTAGCAGAAC CTAATGTACG ATAATTGGGA AGAACGCGGG AAACCGTTTG
 1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

This corresponds to the amino acid sequence [<SEQ ID 322; ORF84-1>] (SEQ ID NO: 322; ORF84-1):

1 MAEICLITGT PGSGKTLKLV SMMANDEMFK PDENGIRRVK FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVVDLYE SAEVHTVNVK
 201 KRKWFYTLV VIVLLIPVTV GLSYKMLSSY GKKQEEPAAG ESAATEQQAV
 251 LPDKTEGEPV NNGNLTADMF VPTLSEKPEP KPIYNGVRQV RTFEYIAGCI
 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGQEVQQS
 351 AQQHSRAQV ATLGGKP*QN LMYDNWEERG KPFEGIGGGV VGSAN*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 (SEQ ID NO: 320) shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) (SEQ ID NO: 324) from strain A of *N. meningitidis*:

orf84 . pep 10 20 30 40 50 60
 MAEICLITGTPGSGKTLKLV SMMANDEMFKPD EKAIRRVFTNIKGLKIPHTYIETDAKK
 orf84a MAEICLITGTPGSGKTLKLV SMMANDEMFKPDENGIRRVFTNIKGLKIPHTYIETDAKK
 45 10 20 30 40 50 60
 orf84 . pep 70 80 90 100 110 120
 LPKSTDEQLSAHDMEYEWIKK PENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG
 orf84a LPKSTDEQLSAHDMEYEWIKK PENIGSIVIVDEAQDVWPARSAGSKI PENVQWLNTHRHQG

-278-

		70	80	90	100	110	120
		130	140	150	160	170	180
	orf84.pep	IDIFVLTQGP	LLDQNLRLTLVR	KHYHIASNKMGM	RTLLEWKICADDP	VKMASSAFSSI	Y
5	orf84a	IDIFVLTQGS	LLDQNLRLTLVR	KHYHIASNKMGM	RTLLEWKICADDP	VKMASSAFSSI	Y
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf84.pep	LDKKVYDLYX	AEVHTVNVKVKR	SKWFYTLPVIV	LLIPVFGVLSY	KMLSSYGKKQE	EPAAQ
10	orf84a	LDKKVYDLYE	SAEHTVNVKVKR	SKWFYTLPVII	LLIPVFGVLSY	KMLSSYGKKQE	EPAAQ
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf84.pep	ESAAATEQQAV	LPDKTEGEPVNN	GNLTADMFPVPT	LSEKXPXSKPI	YNGVRQVRTFE	YIAGCI
15	orf84a	ESAAATEHQAV	FQDKTEGEPVNN	GNLTADMFPVPT	LSEKPESKPI	YNGVRQVRTFE	YIAGCV
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf84.pep	EGGRTGCACYS	HQGTALKEVTE	LMCKDYVKNGL	PFNPYKEESQ	GQEVQQAQQH	SDRAQV
20	orf84a	EGGRTGCTCY	SHQGTALKEIT	KEMCKDYARN	GLPFNPYKEES	QGRDVQQSEQ	HHSDRPQV
		310	320	330	340	350	360
		370	380	390			
	orf84.pep	ATLGGKXPXQ	NLMYDNWEERG	KPFEGIGGGV	VGSANX		
25	orf84a	ATLGGKWPQ	NLMYDNWQERG	KPFEGIGGGV	VGSANX		
		370	380	390			

The complete length ORF84a nucleotide sequence [<SEQ ID 323>] (SEQ ID NO: 323) is:

	1	ATGGCAGAGA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTCAG	GGAAAAACATT
30	51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAAG	CCGGATGAAA
	101	ACGGCATACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGCTT	GAAGATACCG
	151	CACACCTACA	TAGAAACGGA	CGCGAAAAAG	CTGCCGAAAT	CGACAGATGA
	201	GCAGCTTTCG	GCGCATGATA	TGTACGAATG	GATAAAGAAG	CCCGAAAATA
	251	TCGGGTCTAT	TGTCATTGTA	GATGAAGCTC	AAGACGTATG	GCCGGCACGC
35	301	TCGGCAGGTT	CAAAAATCCC	TGAAAATGTC	CAATGGCTGA	ATACGCACAG
	351	ACATCAGGGC	ATTGATATAT	TTGTTTGTAC	TCAAGGCTCT	AAGCTTCTAG
	401	ATCAAAATCT	TAGAACGCTT	GTACGGAAAC	ATTACCACAT	CGCTTCAAAC
	451	AAGATGGGTA	TGCGTACGCT	TTTAGAATGG	AAAATATGCG	CGGACGATCC
	501	CGTAAAAATG	GCATCAAGCG	CATTCTCCAG	TATCTATACA	CTGGATAAAA
40	551	AAGTTTATGA	CTTGATACGAA	TCAGCGGAAG	TTCATACCGT	AAATAAGGTC
	601	AAGCGGTCAA	AATGGTTTTA	TACTCTGCCA	GTAATAATAT	TGCTGATTCC
	651	CGTTTTTGTC	GGCCTGTCCCT	ATAAAATGTT	AAGTAGTTAT	GGAAAAAAAC
	701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CAACAGAACA	TCAGGCAGTA
	751	TTTCAGGATA	AAACAGAAGG	CGAGCCGGTA	AACAACGGTA	ACCTTACCGC
45	801	AGATATGTTT	GTTCCGACAT	TGTCCGAAAA	ACCCGAAAGC	AAGCCGATTT
	851	ATAACGGTGT	AAGGCAGGTA	AGAACCCTTG	AATATATAGC	AGGCTGTGTA
	901	GAAGGCGGAA	GAACCGGATG	CACATGCTAT	TCGCATCAAG	GGACGCATT
	951	GAAAGAAATT	ACAAAGGAAA	TGTGCAAGGA	TTACGCAAGA	AACGGATTGC
	1001	CGTTTAACCC	ATATAAGAAA	GAAAGCCAAAG	GGCGGGATGT	CCAGCAAAGT
50	1051	GAGCAGCACC	ATTCGGACAG	ACCGCAAGTT	GCCACGTTGG	GCGGAAAGCC
	1101	GTGGCAAAAT	CTTATGTATG	ATAATTGGCA	GGAGCGCGGA	AAACCGTTTG
	1151	AAGGAATCGG	CGGGGGCGTG	GTCGGATCGG	CAAACCTGA	

This encodes a protein having amino acid sequence [<SEQ ID 324>] (SEQ ID NO: 324):

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGS KLLDQNLRTL VRKHYHIASN
 151 KGMRTLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNKV
 5 201 KRSKWFTLP VIILLIPVFV GLSYKMLSSY GKKQEEPAAQ ESAATEHQAV
 251 FQDKTEGEPV NNGNLTADMF VPTLSEKPES KPIYNGVRQV RTFEYIAGCV
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPHYKE ESQGRDVQGS
 351 EQHHS DRPQV ATLGKGPWQN LMYDNWQERG KPFEGIGGGV VGSAN*

10 ORF84a (SEQ ID NO: 324) and ORF84-1 (SEQ ID NO: 322) show 95.2% identity in 395 aa overlap:

		10	20	30	40	50	60
	orf84a.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK					
15	orf84-1	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK					
		10	20	30	40	50	60
	orf84a.pep	70	80	90	100	110	120
		LPKSTDEQLSAHDMEYEWIKK PENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
20	orf84-1	LPKSTDEQLSAHDMEYEWIKK PENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
		70	80	90	100	110	120
	orf84a.pep	130	140	150	160	170	180
		IDIFVLTQGSKLLDQNLRTLVRKHYHIASNKGMRTLLEWKICADDPVKMASSAFSSIYT					
25	orf84-1	IDIFVLTQGPKLLDQNLRTLVRKHYHIASNKGMRTLLEWKICADDPVKMASSAFSSIYT					
		130	140	150	160	170	180
	orf84a.pep	190	200	210	220	230	240
		LDKKVYDLYESA EVHTVNKV KRSKWFTLPV IILLIPVFVGLSYKMLSSYGKKQEEPAAQ					
30	orf84-1	LDKKVYDLYESA EVHTVNKV KRSKWFTLPV IIVLLIPVFVGLSYKMLSSYGKKQEEPAAQ					
		190	200	210	220	230	240
	orf84a.pep	250	260	270	280	290	300
		ESAATEHQAVFQDKTEGEPVNNGNLTADMFVPTLSEKPESKPIYNGVRQVRTFEYIAGCV					
35	orf84-1	ESAATEQQA VLPDKTEGEPVNNGNLTADMFVPTLSEKPESKPIYNGVRQVRTFEYIAGCI					
		250	260	270	280	290	300
	orf84a.pep	310	320	330	340	350	360
		EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPHYKEESQGRDVQQSEQHHS DRPQV					
40	orf84-1	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPHYKEESQGEVQQSAQQHSDRAQV					
		310	320	330	340	350	360
	orf84a.pep	370	380	390			
		ATLGKGPWQN LMYDNWQERG KPFEGIGGGV VGSANX					
45	orf84-1	ATLGKGPXQN LMYDNWEERG KPFEGIGGGV VGSANX					
		370	380	390			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 (SEQ ID NO: 320) shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) (SEQ ID NO: 326) from *N. gonorrhoeae*:

5	orf84.pep	MAEICLITGTPGSGKTLKMVSMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
	orf84ng	MAEICLITGTPGSGKTLKMVSMANDEMFKPDENGVRKVFTNIKGLKIPHTHIETDAKK	60
	orf84.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
	orf84ng	LPKSTDEQLSAHDMYEWIKKPENVGAIIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
10	orf84.pep	IDIFVLTQGPVKLLDQNLRLTLVRKHYHIASNKMGMRTLLEWKICADDPVKMASSAFSSIYT	180
	orf84ng	IDIFVLTQGPVKLLDQNLRLTLVRKHYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIYT	180
	orf84.pep	LDKKVYDLYXXAEVHTVNKVKRSKWFTLPVIVLLIPVFGVLSYKMLSSYGKKQEPPAAQ	240
	orf84ng	LDKKVYDLYESAEIHTVNKVKRSKWFTLPVILLIPLFVGLSYKMLGSYGKKQEPPAAQ	240
15	orf84.pep	ESAAEQQAVLPDKTEGEPVNNGNLTADMVFPVTLSEKXPXSKPIYNGVRQVRTFEYIAGCI	300
	orf84ng	ESAAEQQAVLPDKTEGESVNNGNLTADMVFPVTLPEKPESKPIYNGVRQVRTFEYIAGCI	300
20	orf84.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSRAQV	360
	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSRAQV	360
	orf84.pep	ATLGGKXPQNLMYDNWEERGKPFEGIGGGVVG SAN 395	
	orf84ng	ATLGGKPPQNLMYDNWEERGKPFEGIGGGVVG SAN 395	

25 The complete length ORF84ng nucleotide sequence [<SEQ ID 325>] (SEQ ID NO: 325) is:

30	1	ATGGCAGAAA	TCTGTTTGAT	AACCGGCACG	CCCGTTTCAG	GGAAAACATT
	51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAG	CCAGATGAAA
	101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAGATACCG
	151	CACACCCACA	TAGAAACAGA	CGCAAAGAAG	CTGCCGAAAT	CAACCGATGA
	201	ACAGCTTTCG	GCGCATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg
35	251	tcggcgCAAT	CGTTATTGTC	GATGAGGCGC	AAGACGTATG	GCCCGCACGC
	301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG
	351	GCATCAGGGC	ATAGATATAT	TTGTATTGAC	ACAAGGTCCT	AAACTCTTAG
	401	ATCAGAACTT	GCGAACATTG	GTTAAAAGAC	ATTACCACAT	TGCGGCCAAC
	451	AAAATGGGTT	TGCGTACCCT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC
40	501	GGTAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA
	551	AAGTTTATGA	CTTGTACGAA	TCCGCAGAAA	TTCAACACGGT	AAACAAAGTC
	601	AAGCGTTCAA	AATGGTTTGA	TGCATTGCCC	GTCATCATAT	TATTGATTCC
	651	GCTATTTGTC	GGTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAAC
	701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CAACAGAACAA	GCAGGCAGTA
45	751	CTTCCGATA	AAACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC
	801	AGATATGTTT	GTTCCGACAT	TGCCCCGAAA	ACCCGAAAGC	AAGCCGATTT
	851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
	901	GAAGGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT
	951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC
	1001	CGTTTTAACCC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
	1051	GCGCAGCAAC	ATTTCGACAG	GGCGCAAGTT	GCCACCTTGG	GCGGAAAGCC
	1101	GCAGCAGAAC	CTAATGTACG	ACAATTGGGA	AGAACGCGGG	AAACCGTTTG
	1151	AAGGAATCGG	CGGGGCGGTG	GTCGGATCGG	CAAACCTGA	

This encodes a protein having amino acid sequence [<SEQ ID 326>] (SEQ ID NO: 326):

```
1  MAEICLITGT PGSGKTLKLV SMMADEMFK PDENGVRKV FTNIKGLKIP
51 HTHIETDAKK LPKSTDEQLS AHDMEYEWIK PENVGAIVIV DEAQDVWPAR
101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VKRHYHIAAN
151 KMGLRTLLEW KVCADDPVKM ASSAFSSIYT LDKKVYDLYE SAEIHTVNKV
201 KRSKWFYALP VIILLIPLFV GLSYKMLGSY GKKQEPPAAQ ESAATEQQAV
251 LPDKTEGESV NNGNLTADMV VPTLPEKPES KPIYNGVRQV RTFEYIAGCI
301 EGGRTGCTCY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGQEVQQS
351 AQQHSDRAQV ATLGKPKQON LMYDNWEERG KPFEGIGGGV VGSAN*
```

ORF84ng (SEQ ID NO: 326) and ORF84-1 (SEQ ID NO: 322) show 95.4% identity in 395 aa overlap:

```
10      20      30      40      50      60
orf84-1.pep MAEICLITGT PGSGKTLKLV SMMADEMFK PDENGVRKV FTNIKGLKIP HTHIETDAKK
15      |||||
orf84ng      MAEICLITGT PGSGKTLKLV SMMADEMFK PDENGVRKV FTNIKGLKIP HTHIETDAKK
10      20      30      40      50      60

70      80      90      100     110     120
orf84-1.pep LPKSTDEQLS AHDMEYEWIK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
20      |||||
orf84ng      LPKSTDEQLS AHDMEYEWIK PENVGAIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
70      80      90      100     110     120

130     140     150     160     170     180
orf84-1.pep IDIFVLTQGP KLLDQNLRTL VKRHYHIA SNKMGMR TLEWKICADD PVKMASSAFSSIYT
25      |||||
orf84ng      IDIFVLTQGP KLLDQNLRTL VKRHYHIA AANKMGLR TLEWKVCADD PVKMASSAFSSIYT
130     140     150     160     170     180

190     200     210     220     230     240
orf84-1.pep LDKKVYDLYE SAEVHTVNKV KRSKWFYTL PVIVLLIP VFVGLSYKMLSSY GKKQEPPAAQ
30      |||||
orf84ng      LDKKVYDLYE SAEIHTVNKV KRSKWFYALP VIILLIPL FVGLSYKMLGSY GKKQEPPAAQ
190     200     210     220     230     240

250     260     270     280     290     300
orf84-1.pep ESAATEQQAV LPDKTEGEPV NNGNLTADM FVPTLSEK PESKPIYNG VRQVRTFEYIAGCI
35      |||||
orf84ng      ESAATEQQAV LPDKTEGESV NNGNLTADM FVPTLPEK PESKPIYNG VRQVRTFEYIAGCI
250     260     270     280     290     300

310     320     330     340     350     360
orf84-1.pep EGGRTGCAC YSHQGTALKEV TELMCKDYVK NGLPFNPY KEESQGQEV QQSAQQHSDRAQV
40      |||||
orf84ng      EGGRTGCTCY SHQGTALKEV TELMCKDYVK NGLPFNPY KEESQGQEV QQSAQQHSDRAQV
310     320     330     340     350     360

370     380     390
orf84-1.pep ATLGKPKXQ NLMYDNWEERG KPFEGIGGGV VGSANX
45      |||||
orf84ng      ATLGKPKQON LMYDNWEERG KPFEGIGGGV VGSANX
370     380     390
```

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 39

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 327>] (SEQ ID NO: 327):

```

1  GTGGTTTTC TGAATGCCGA CAACGGGATA TTGGTTCAGG ACTTGCCCTT
51  TGAAGTCAAA CTGAAAAAAT TCCATATCGA TTTTACAAT ACGGGTATGC
10  CGCGTGATTT CGCCAGCGAT ATTGAAGTGA CGGACAAGGC AACCGGTGAG
151  AAACCTCGAGC GCACCATCCG CGTGAACCAT CCTTTGACCT TGCACGGCAT
201  CACGATTTAT CAGGCGAGTT TTGCCGACGG CGGTTCGGAT TTGACATTCA
251  AGGCGTGGAA TTTGGGTGAT GCTTCGCGCG AGCCTGTCGT GTTGAAGGCA
301  ACATCCATAC ACCAGTTTCC GTTGGAATTT GGCAAACACA AATATCGTCT
15  351  TGAGTTCGAT CAGTTCACCT CTATGAATGT GGAGGACATG AGCGAGGGCG
401  CGGAACGGGA AAAAAGCCTG AAATCCACGC TGCCCGATGT CCGCGCCGTT
451  ACTCAGGAAG GTCACAAATA CACCAAT... ..TACCG
501  TATCCGTGAT GCGCCAGGCC AGGCGGTCGA ATATAAAAAC TATATGCTGC
551  CGGTTTTGCA GGAACAGGAT TATTTTGGGA TTACCGGCAC GCGCAGCGC.
20  601  TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
651  AGCGGACACC TTTATGGCAT TGCGTGAGTT TTTGAAAGAT GGGGAAGGGC
701  GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
751  GAACAATTCA TGCTGGCTGC GGAAAACACG CTGAACATCT TTGCACAAAA
801  AGGCTATTTG GGATTGGACG AATTATTATC GTCCAATATC CCGAAAGAGC
25  851  AGCAGGATAA GATGCAGGGC TATTTCTACG AAATGCTTTA CGGCGTGATG
901  AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
951  GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTCGGG TTTGCAGATG ACCCGTTCCT C.GGTCCGCT
30  1101 TTTGGTCTAT CTC...
```

This corresponds to the amino acid sequence [<SEQ ID 328; ORF88>] (SEQ ID NO: 328; ORF88):

```

1  MVFLNADNGI LVQDLPFVEK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
35  51  KLERTIRVNH PLTLHGITYI QASFADGGSD LTFKAWN LGD ASREPVLKA
101  TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRAV
151  TQEGHKYTNX XXXXXYRIRD APGQAVEYKN YMLPVLQEQD YFWITGTRSX
201  LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRXVAD ATKGAPAEIR
251  EQFM LAAENT LNIFAQKGYL GLDEFITSNI PKEQQDKMQG YFYEMLYGVM
40  301  NAALDETXTR YGLPEWQQDE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
351  SEVRSSGLQM TRSXGPLLVY L...
```

Further work revealed the complete nucleotide sequence [<SEQ ID 329>] (SEQ ID NO: 329):

```

1  ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTCGC
45  51  TTTTTCAGC TCCATGCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA
```

5
10
15
20
25
30
35
40

```

101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTTCCTGG
301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA
401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTGCCTG GCGGGTTGA
551 TAGACAGTAA CCTGCTGTTG AAACCTGGGTG TGCTGACCGG TCGGATTGTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCGTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCGGA TGTGTTTTC CTGAATGCCG ACAACGGGAT ATTGGTTCAG
751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAGG
851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCTG
1001 TGTTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTCTG AATATAAAAA
1251 CTATATGCTG CCGGTTTTCG AGGAACAGGA TTATTTTGG ATTACCGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGGAAGGG CGCAAACGTC TGTTTGCCGA CGCAACCAA GCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTCC
1801 CCGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GGCGTGGGTA TTGTTTTCAG
1901 ACGGCAAAAT CCGTTTTCGCC ATGTCTTCGG CCCGCAGCGA ACGGGATTTG
1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA

```

This corresponds to the amino acid sequence [<SEQ ID 330; ORF88-1>] (SEQ ID NO: 330; ORF88-1):

45
50
55

```

1  MSKSRSPPL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
51  YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SLLLDVKIAP EVAKRYLEVQ GFQGKTINRE
151 DGSVLIAAKK GTMNKWGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITIYQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PVLQEODYFW ITGTRSLGQQ QYRWLRIPLD
451 KQLKADTFMA LREFLKDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI
501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGMNAA LDETIRRYGL
551 PEWQQDEARN RFLHSHMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
601 PGALLVYLG S VLLVLGTVLM FYVREKRAWV LFS DGKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF88 (SEQ ID NO: 328) shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) (SEQ ID NO: 332) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF88a nucleotide sequence [<SEQ ID 331>] (SEQ ID NO: 331) is:

45

1	ATGAGTAAAT	CCCGTAGATC	TCCCCACTT	CTTTCCTGTC	CGTGGTTTCGC
51	TTTTTTCAGC	TCCATGCGCT	TTGCGGTTCGC	TTTGCTCAGT	CTGCTGGGTA
101	TTGCATCGGT	TATCGGTACG	GTGTTGCAGC	AAAACCAGCC	GCAGACGGAT

5
10
15
20
25
30
35

```

151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTATATC ATGATGTTTT
251 TGGTGGTTTT TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTTCTGG
301 CGCGAAATGA AGTCTTTTTC GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA
401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTGCGCTG GGCGGGTTGA
551 TAGACAGTAA CCTGCTGTTG AAAC TGCGTACCGG TCGGATTGTT
601 CCGGACAATC AGCGGTTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCGGA TGTGGTTTTT CTGAATGCCG ACAACGGGAT ATTGGTTCAG
751 GACTTGCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTTACAA
801 TACGGGTATG CCGCGCGATT TTGCCAGTGA TATTGAAGTA ACGGATAAGG
851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTGCGCGACG GCGGTTCCGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCTG
1001 TGTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTTACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTA AAAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TCGGCGAGG CAGGCGGTCTG AATATAAAAA
1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GGCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTCAGAT GACCCGTTCC
1801 CCGGGTGC GC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GGCGTGGGTA TTGTTTTAG
1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCGCAGCGA ACGGGATTG
1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA

```

40 This encodes a protein having amino acid sequence [SEQ ID 332] (SEQ ID NO: 332):

45
50
55

```

1 MSKSRRSPPL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVQ GFQKGTINRE
151 DGSVLIAAKK GTMNKWDYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITYQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKTYNIGPS
401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSLQOQ QYRWLRIPLD
451 KQLKADTFMA LREFLKDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI
501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVMNAA LDETIRRYGL
551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFSDGKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*

```

ORF88a (SEQ ID NO: 332) and ORF88-1 (SEQ ID NO: 330) 100.0% identity in 671 aa overlap:

	orf88-1	MSKSRRSPPLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLVKFGSFWA	60
	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
5	orf88-1	QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKGYIFAHVALIVICL	180
	orf88-1	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKGYIFAHVALIVICL	180
10	orf88a.pep	GGLIDSNNLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
	orf88-1	GGLIDSNNLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
	orf88a.pep	LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88-1	LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
15	orf88a.pep	LHGITIYQASFADGGSDLTFAWNLGDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88-1	LHGITIYQASFADGGSDLTFAWNLGDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
20	orf88-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88a.pep	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
25	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88-1	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88-1	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
30	orf88a.pep	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88-1	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88a.pep	LQRLGKDLNHD	672
35	orf88-1	LQRLGKDLNHD	672

Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 (SEQ ID NO: 328) shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) (SEQ ID NO: 334) from *N. gonorrhoeae*:

40	orf88.pep	MVFLNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88ng	MVFLNADNGMLVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60

	orf88.pep	PLTLHGITIYQASFADGGSDLTFKAWNLDASREPVLKATSIHQFPLEIGKHKYRLEFD	120
	orf88ng	PLTLHGITIYQASFADGGSDLTFKAWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFD	120
5	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRVAVTQEGHKYTNXXXXXXYRIRDAPGQAVEYKN	180
	orf88ng	QFTSMNVEDMSEGAEREKSLKSTLNDVRVAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
	orf88.pep	YMLPVLQEODYFWITGTRSXLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRXVAD	240
	orf88ng	YMLPILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVAD	240
10	orf88.pep	ATKGAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM	300
	orf88ng	ATKDAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVM	300
	orf88.pep	NAALDETXTRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
15	orf88ng	NAALDETIIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88.pep	TRSXGPLLVL	371
	orf88ng	TRSPGALLVYLGSVLLVLGTVMFYVPKKRAWVLFNSXKIRFAMSSARSERDLQKEFPKH	420

- 20 An ORF88ng nucleotide sequence [<SEQ ID 333>] (SEQ ID NO: 333) was predicted to encode a protein having amino acid sequence [<SEQ ID 334>] (SEQ ID NO: 334):

	1	MVFLNADNGM	LVQDLPFVEVK	LKKFHIDFYN	TGMPRDFASD	IEVTDKATGE
	51	KLERTIRVNH	PLTLHGITIY	QASFADGGSD	LTFKAWNLRD	ASREPVLKA
25	101	TSIHQFPLEI	GKHKYRLEFD	QFTSMNVEDM	SEGAEREKSL	KSTLNDVRVAV
	151	TQEGKKYTN	GPSIVYRIRD	AAGQAVEYKN	YMLPILQDKD	YFWLTGTRSG
	201	LQQQYRWLR	IPLDKQLKADT	FMALREFLKD	GEGRKRLVAD	ATKDAPAEIR
	251	EQFMLAAENT	LNIFAQKGYL	GLDEFITSNI	PKGQDKMQG	YFYEMLYGVM
	301	NAALDETIIR	YGLPEWQQDE	ARNRFLHSM	DAYTGLTEYP	APMLLQLDGF
	351	SEVRSSGLQM	TRSPGALLVY	LGSVLLVLGT	VFMFYVPKKR	AWVLFNSXKI
30	401	RFAMSSARSE	RDLQKEFPKH	VESLQRLGKD	LNHD*	

Further work revealed the complete gonococcal DNA sequence [<SEQ ID 335>] (SEQ ID NO: 335):

	1	ATGAGTAAAT	CCCGTATATC	TCCCACACTT	CTTCCCCTGC	CGTGGTTCGC
35	51	TTTTTTCAGC	TCCATGCGCT	TTGCCGTCGC	TTTGCTCAGT	CTGCTGGGTA
	101	TTGCATCGGT	TATCGGCACG	GTGTTACAGC	AAAACCAGCC	GCAGACGGAT
	151	TATTTGGTCA	AATTCGAC	GTTTTGGACT	CGGATTTTGT	ATTTTTTGGG
	201	TTTGTATGAT	GTCTATGCTT	CGGCATGGTT	TGTCGTTATC	ATGATGTTTC
	251	TGGTGGTTTC	TACCACTTTC	TGTTTAATCC	GTAACGTTCC	GCCGTTTTGG
40	301	CGCGAAATGA	AGTCTTTCCG	GGAAAAGGTT	AAAGAAAAAT	CTCTGGCGGC
	351	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAA	AATTGCCCCC	GAAGTTGCCA
	401	AACGTTATCT	GGAGGTGCGG	GGTTTTCAGG	GAAAAACCGT	CAGCCGTGAG
	451	GACGGGTCGG	TTCTGATTGC	CGCCAAAAAA	GGCAaatga	acaaATGGGG
	501	CTATATCTTT	GCcCaagtag	ctTTGATTGT	CATTTGCCTG	GGCGGGTTGA
45	551	TAGACAGTAA	CCTGCTGCTG	AAGCTGGGTA	TGCTGGCCGG	TCGGATTGTT
	601	CCGGACAATC	AGGCGGTTTA	TGCCAAGGAT	TTCAAGCCCG	AAAGTATTTT
	651	GGGTGCGTCC	AATCTCTCAT	TTAGGGGCAA	CGTCAATATT	TCCGAGGGGC
	701	AAAGTGC	GGA TGTGGTTTTC	CTGAATGCCG	ACAACGGGAT	GTTGGTTCAG
	751	GACTTGCCCTT	TTGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTACAA

5	801	TACGGGTATG	CCGCGCGATT	TTGCCAGCGA	TATTGAAGTA	ACGGACAAGG
	851	CAACCGGTGA	GAAACTCGAG	CGCACCATCC	GCGTGAACCA	TCCTTTGACC
	901	TTGCACGCA	TCACGATTTA	TCAGGCGAGT	TTTGCCGACG	GCGGTTCTGGA
	951	TTTGACATTC	AAGGCGTGGA	ATTTGAGGGA	TGCTTCGCGC	GAACCTGTCTG
	1001	TGTTGAAGGC	AACCTCCATA	CACCACTTTC	CGTTGGAAAT	CGCGCAACATA
10	1051	AAATATCGTC	TTGAGTTCGA	TCAGTTCACT	TCTATGAATG	TGGAGGACAT
	1101	GAGCGAGGGT	GCGGAACGGG	AAAAAAGCCT	GAAATCCACT	CTGAACGATG
	1151	TCCGCGCCGT	TACTCAGGAA	GGTAAAAAAT	ACACCAATAT	CGGCCCTTCC
	1201	ATCGTGTACC	GCATCCGTGA	TGcggCAGGG	CAGGCGGTCTG	AATATAAAAA
	1251	CTATATGCTG	CCGATTTTGC	AGGACAAAAG	TTATTTTGTG	CTGACCCGGAC
15	1301	CGCGCAGCGG	CTTGCAGCAG	CAATACCGCT	GGCTGCGTAT	CCCTTGGCA
	1351	AAGCAGTTGA	AAGCGGACAC	CTTTATGGCA	TGCGTGTAGT	TTTTGAAAGA
	1401	TGGGGAAGGG	CGCAAACGTC	TGGTTGCCGA	CGCAACCAAA	GACGCACCTG
	1451	CCGAAATCCG	CGAACAATTC	ATGCTGGCTG	CGGAAAACAC	GCTGAATATC
	1501	TTTGCGCAAA	AAGGCTATTT	GGGATTGGAC	GAATTTATTA	CGTCCAATAT
20	1551	CCCGAAAGGG	CAGCAGGATA	AGATGCAGGG	CTATTTCTAC	GAAATGCTTT
	1601	ACGCGGTGAT	GAACGCTGCT	TTGGATGAAA	CCATACGCCG	GATACGGCTTG
	1651	CCCGAATGGC	AGCAGGATGA	AGCGCGGAAC	CGTTTCTCTG	TGCACAGTAT
	1701	GGATGCCTAT	ACGGGGCTGA	CGGAATATCC	CGCGCCTATG	CTGCTCCAGC
	1751	TTGACGGGTT	TTCCGAGGTG	CGTTCCTCAG	GTTTGCAGAT	GACCCGTTCTG
25	1801	CCGGGTGCGC	TTTTTGGTCTA	Tctcggctcg	gtattggttg	TTTTGGgtac
	1851	ggtaTttatg	tTTTATGTGC	GCGAAAAACG	GGCGTGGgta	tTGTTTTCTag
	1901	aCGGCCAAAA	CCGTTTTGCT	ATGtCTTcgg	CCcgagcga	ACGGGATTTG
	1951	cAGAaggaat	TTCCAAAAACA	CGtcgAGAGC	CTGCAACggc	tcggcaaggA
	2001	CttgaaTCAT	GACTga			

This corresponds to the amino acid sequence [<SEQ ID 336; ORF88ng-1>] (SEQ ID NO: 336; ORF88ng-1):

30	1	MSKSRI\$PTL	LSRPWF\$FFS	SMRFAVALLS	LLGIASVIGT	VLQONQPQTD
	51	YLVKFGPFWT	RIFDFLGLYD	VYASAWFVVI	MMFLVVSTSL	CLIRNVPPFW
	101	REMKSFREKV	KEKSLAAMRH	SSLLDVKIAP	EVAKRYLEVR	GFQKTVSRE
	151	DGSVLIAAKK	GTMNKWGYIF	AQVALIVICL	GGLIDSNLLL	KLGMLAGRIV
	201	PDNQAVYAKD	FKPESILGAS	NLSFRGNVNI	SEGO\$ADVV\$	LNADNGMLVQ
35	251	LDPFVEVLK\$	PHIDFYNTGM	PRDFASDIEV	TDKATGEKLE	RTIRVN\$PLT
	301	LHGITYIQAS	FADGGS\$DLT\$	KAWNLRDASR	EPVVLKATS\$	HQ\$P\$LEIGHK
	351	KYRLEFDQFT	SMNVEDMSEG	AEREKSLKST	LNDVRAVTQE	GKKYTNIGPS
	401	IVYRIRDAAG	QAVEYKNYML	PILQDKDYFW	LTGTR\$GLQ\$	QYRWLRIPLD
	451	KQLKADTFMA	LREFLKDGE\$	RKRLVADATK	DAPAEIREQF	MLAAENTLNI
40	501	FAQKG\$YGLD	EFITSNIPKG	QQDKMQGYFY	EMLYGVMNAA	LDETIRRYGL
	551	PEWQQDEARN	RFLH\$MSDAY	TGLTEYPAPM	LLQLDGFSEV	RSSGLQMT\$S
	601	PGALLVYLGS	VLLVLGTVFM	FYVREKRAWV	LFSDGKIRFA	MSSARSERDL
	651	OKEFFPKHVES	LQRLGKDLNH	D*		

ORF88ng-1 (SEQ ID NO: 336) and ORF88-1 (SEQ ID NO: 330) show 97.0% identity in 671 aa
45 overlap:

	orf88-1.pep	MSKSRRSPPLLSRPWF AFFSSMRFAVALLSLLGIASVIGTVLQONQPQT DYLVKFGSFWA	60
	orf88ng-1	MSKSRI SPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQT DYLVKGFPFWT	60
50	orf88-1.pep	QIFGFLGLDYVYASAWFVVIMMFLV VSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf88ng-1	RIFDFLGLDYVYASAWFVVIMMFLV VSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf88-1.pep	SSLDDVKIAPEVAKRYLEVQGFGQKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL	180

	orf88ng-1	SSLLDVKIAPEVAKRYLEVRGFQGKTVSREDGSVLIAAKKGTMNKWKGYIFAQVALIVICL	180
	orf88-1.pep	GGLIDSNLLKLKGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
5	orf88ng-1	GGLIDSNLLKLKGLMAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF	240
	orf88-1.pep	LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88ng-1	LNADNGMLVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
10	orf88-1.pep	LHGITIYQASFADGGSDLTfKAWN LGDASREP VVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88ng-1	LHGITIYQASFADGGSDLTfKAWN LR DASREP VVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88-1.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88ng-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
15	orf88-1.pep	PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88ng-1	PILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
20	orf88ng-1	DAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVMNAA	540
	orf88-1.pep	LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88ng-1	LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
25	orf88-1.pep	PGALLVYLGSVLLVLGTVLMFVYREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88ng-1	PGALLVYLGSVLLVLGTVFMFVYREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88-1.pep	LQRLGKDLNHD	671
30	orf88ng-1	LQRLGKDLNHD	671

Furthermore, ORG88ng-1 (SEQ ID NO: 336) shows homology with a hypothetical protein (SEQ ID NO: 1134) from *Aquifex aeolicus*:

35	gi 2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537 Score = 94.4 bits (231), Expect = 2e-18 Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)
	Query: 16 FAFFSSMRFAVALLSLLGIASVIG-TVLQONQPQTDYLVKFGPFWTRIFDFGLGYDVYAS 74 + F +S++ A+ ++ +LGI S++G T ++QNQ YL +FG L L DV+ S
	Sbjct: 80 YDFLASLKLAI FIMLVLGILSMLGSTYIKQNQSF EWYLDQFGYDVGIWIWKLWLNDFVHS 139
40	Query: 75 AWFVIMMFLVVSLSLCLIRNVPPFWREMKSFREKVKEKSLAAMRHSSLLDVKIAPEVAK 134 ++++ ++ L V+ C I+ +P W++ S +E++ + A +H + VKI P+ K
	Sbjct: 140 WYYILFIVLLAVNLIFCSIKRLPRVWKQAFS-KERILKLDEHAEKHLKPITVKI-PDKDK 197
	Query: 135 --RYLEVRGFQGKTVSREDGSVLIAAKKGT MNKWKGYIFAQVALIVICLGG LIDSNLLKL 192 ++L +GF+ V E + + A+KG ++ G +AL+VI G LID
	Sbjct: 198 VLKFLKKGFK-VFVEEENKLYVFAEKGRFSRLGVYITHIALLVIMAGALID----- 249

Query: 193 GMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVFLNADNGMLVQDL 252
 +I+G RG++ ++EG + DV+ + A+ L
 Sbjct: 250 -----AIVGV-----RGSLLVAEGDTNDVMLVGAE--QKPYKL 280

Query: 253 PFEVKLKKFHIDFY---NTGMPRDFA-----SDIEVTDKATGEKLER--TIRVNHPLT 300
 PF V L F I Y N + + FA SDIE+ + G K+E T++VN P
 Sbjct: 281 PFAVHLIDFRIKTYAEENPNVDKRFAQAVSSYESDIEIIN--GGKVEAKGTVKVNEPFD 337

Query: 301 LHGITIYQASFA--DGGSDLTFKAWNLRDASREP 332
 ++QA++ DG S + + + A +P
 Sbjct: 338 FGRYRLFQATYGILDGTSGMGVIVVDRKKAHEDP 371

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 40

- 15 The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 337>] (SEQ ID NO: 337):

1 ATGATGAGTA ATAmAATGGm ACAAAAAGGG TTTACATTGA TTGmGmTGAT
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
 101 ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
 151 GyCGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
 201 CGATAATCAG ACCATCGAGA ACAAAGTGG AATATTTGTC TCAGGCTATA
 251 AGATGAATCC GAAAATTGCC AAAAAaTATA GTGTTTCGGT AAAGTTTGTC
 301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGCGGTTT CGAAGGCGGG
 351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
 25 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

This corresponds to the amino acid sequence [<SEQ ID 338; ORF89>] (SEQ ID NO: 338; ORF89):

1 MMSNXMXQKG FTLIXXMIVV AILGIISVIA IPSYXSYIEK GYQSOLYTEM
 51 XGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKPV
 101 DKEKSRAAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLS
 151 DVGCEAFSNR KK*

- 35 Further work revealed the complete nucleotide sequence [<SEQ ID 339>] (SEQ ID NO: 339):

1 ATGATGAGTA ATAAAAATGGA ACAAAAAGGG TTTACATTGA TTGAGATGAT
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
 101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
 151 GTCGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
 201 CGATAATCAG ACCATCGAGA ACAAAGTGG AATATTTGTC TCAGGCTATA
 251 AGATGAATCC GAAAATTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC
 301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGCGGTTT CGAAGGCGGG
 351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA

5 1 MMSNKMEQKG FTLIEMMIVV AILGIISVIA IPSYQSYIEK GYQSQLYTEM
 51 VGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNP KIA KKYSVSVK FV
 101 DKEKSRAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLSS
 151 DVGCEAFSNR KK*

15

orf89	8	QKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQFILKNPL-	66
		QKGFTLI MIV+AI+GI++ +A+P+Y Y + S+ G + ++ L + +	
Pile	5	QKGFTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGIW	64
orf89	67	-DDNQTIENKLEIFVSGYKMPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGYYTLVSW	125
		DN + +G + KI KY SV + GV K G LS+W	
Pile	65	PKDNTS-----AGVASSDKIKGYVQSVTVAKGVVTAEMASTGVNKEIQGKKLSLW	115

The complete length ORF89a nucleotide sequence [<SEQ ID 341>] (SEQ ID NO: 341) is:

5
10
15
20
25
30
35
40
45

```

1  ATGATGAGTA ATAAAATGGA ACAAAAAGGG TTTACATTGA TTGNGANGNT
51 NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT
101 ATCNNAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAATATTTT CAAACAGTNT ATTTTGAAAA ATCCCCTGGA
201 CGATAATCAG ACCATCAAGA GCAAACAGT AATATTTGTC TCAGGCTATA
251 AGATGAATCC GAAAATTGCC GAAAAATATA ATGTTTCGGT GCATTTTGTG
301 AATGAGGAAA AACCNAGGGC ATACAGCTTG GTCGGCGTTC CAAAGACGGG
351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGGAGAC CTTGTCCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

```

This encodes a protein having amino acid sequence [<SEQ ID 342>] (SEQ ID NO: 342):

15
20
25
30
35
40
45

```

1  MMSNKMEQKG FTLIXXXXXX AIXXXSVIX XXXYXSYIEK GYQSQLYTEM
51 VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHVF
101 NEEKPRAYSL VGVPKTGTGY TLSVWMNSVG DGYKCRDAAS ARAHLETLS
151 DVGCEAFSNR KK*

```

ORF89a (SEQ ID NO: 342) and ORF89-1 (SEQ ID NO: 340) show 83.3% identity in 162 aa overlap:

20
25
30
35

```

      10      20      30      40      50      60
orf89a.pep MMSNKMEQKGFTLIXXXXXXAIXXXSVIXXXYXSYIEKGYQSQLYTEMVGINNISKQX
          |||||
orf89-1    MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF
          10      20      30      40      50      60

      70      80      90      100     110     120
orf89a.pep ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHVFVNEEKPRAYSLVGVPKTGTGY
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf89-1    ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY
          70      80      90      100     110     120

      130     140     150     160
orf89a.pep TLSVWMNSVG DGYKCRDAAS ARAHLETLS DVGCEAFSNRKKX
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf89-1    TLSVWMNSVG DGYKCRDAASA QAHLETLS DVGCEAFSNRKKX
          130     140     150     160

```

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF89 (SEQ ID NO: 338) shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) (SEQ ID NO: 344) from *N. gonorrhoeae*:

40

```

orf89      MMSNXMXQKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQF 60
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf89ng    MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF 60

orf89      ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY 120
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf89ng    ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY 120

orf89      TLSVWMNSVG DGYKCRDAASA QAHLETLS DVGCEAFSNRKK 162

```

orf89ng |||||:||||:||||:|||||
 TLNVWMNSVGDGYKCRDATSAQAYSDTLNADSGCEAFSNRKK 162

The complete length ORF89ng nucleotide sequence [<SEQ ID 343>] (SEQ ID NO: 343) is:

```

5      1  aTGATGAGCA ATAAATGGA AAAAAAGGG TTTACATTGA TTGAGATGAT
      51  GATAGTTGTC ACGATACTCG GCATCATCAG CGTCATTGCC ATACCTTCTT
     101  ATCAGAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
     151  GTCGGTATCA ACAATGTTCT CAAACAGTTT ATTTTGAAAA ATCCCAGGA
     201  CGATAATGAT ACCCTCAAGA GCAAACTGAA AATATTTGTC TCAGGCTATA
    10  251  AGATGAATCC GAAAAttgCC AAAAAATATA GTGTTTCGGt aaggtttGTC
      301  gatGCGGAAA AACCAAGGGC ATACAGGTTG GTCGCGGTC CGAACGCGGG
      351  GACGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGATACA
     401  AATGCCGTGA TGCCACTTCT GCCCAGGCCT ATTCGGACAC CTTGTCCGCA
     451  GATAGCGGCT GTGAAGCTTT CTCTAATCGT AAAAAATAG
  
```

This encodes a protein having amino acid sequence (SEQ ID NO: 344):

```

      1  MMSNKMEQKG FTLIEMMIVV TILGIISVIA IPSYQSYIEK GYQSQLYTEM
     51  VGINNVLKQF ILKNPQDDND TLKSKLKIFV SGYKMNPKIA KKYSVSVRFV
    101  DAEKPRAYRL VGVPNAGTGY TLNVWMNSVG DGYKCRDATS AQAYSDTLN
    151  DSGCEAFSNR KK*
  
```

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng (SEQ ID NO: 344) and ORF89-1 (SEQ ID NO: 340) show 88.3% identity in 162 aa overlap:

```

25      10      20      30      40      50      60
    orf89-1.pep  MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF
    orf89ng      MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF
                10      20      30      40      50      60

    30      70      80      90      100     110     120
    orf89-1.pep  ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAVRLVGVPKAGTGY
    orf89ng      ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY
                70      80      90      100     110     120

    35      130     140     150     160
    orf89-1.pep  TLNVWMNSVGDGYKCRDAASAQAHLTLSSDVGCEAFSNRKKX
    orf89ng      TLNVWMNSVGDGYKCRDATSAQAYSDTLNADSGCEAFSNRKKX
                130     140     150     160
  
```

Based on this analysis, including the gonococcal motifs and the homology with the known Pile protein (SEQ ID NO: 1135), it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (SEQ ID NO: 340) (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that ORF89-1 (SEQ ID NO: 340) is a surface-exposed protein, and that it is a useful immunogen.

Example 41

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 345] (SEQ ID NO: 345):

```

10      1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
      51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
     101  ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
     151  CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTCGATT TCCAACGTAT
     201  GACCGCATTG GCGGTCGGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
     15  251  AACAAGCGTT GGCCn.AGAA TTTCAACCC...
```

This corresponds to the amino acid sequence [SEQ ID 346; ORF91] (SEQ ID NO: 346; ORF91):

```

20      1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
     51  RQKAEAYAIP YDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...
```

Further work revealed the complete nucleotide sequence [SEQ ID 347] (SEQ ID NO: 347):

```

25      1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
     51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
     101  ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
     151  CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTCGATT TCCAACGTAT
     201  GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
     251  AACAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
     301  GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
     351  CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
     401  TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
     451  GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
     501  CGTGTAACGC AACCATTTCG GCGAAATTAT CAAAGCGAAA GGCGTGACG
     551  GACTGATTGC CGAGTTGAAA GCCAAAAACG GCGGCAAATA A
```

This corresponds to the amino acid sequence [SEQ ID 348; ORF91-1] (SEQ ID NO: 348; ORF91-1):

```

40      1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
     51  RQKAEAYAIP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
    101  GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
```

151 GKYRTYNVAI EGASLVTVYR NQFGEIIKAK GVDGLIAELK AKNGGK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 5 ORF91 (SEQ ID NO: 346) shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) (SEQ ID NO: 350) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf91.pep	MKKSS	LISALGIGILSIGMAFAAPADAVSQIRQ	NATQVLSILKNGDANTARQKAEAYAIP			
10	orf91a	MKKSSF	LISALGIGILSIGMAFAAPADAVNQIRQ	NATQVLSILKSGDANTARQKAEAYAIP			
		10	20	30	40	50	60
		70	80	90			
	orf91.pep	YFDFQ	RMTALAVGNPWXTXSDXQKQALAXEFQP				
15	orf91a	YFDFQ	RMTALAVGNPWRTASDAQKQALAKEFQTL	LIRTYSGTMLKLNANVNVKDNPIVN			
		70	80	90	100	110	120
	orf91a	KGGKE	IIVRAEVGVPQKPVNMDFTTYQSGGKYR	TYNVAIEGASLVTVYRNQFGEI	I	KAK	
		130	140	150	160	170	180

- 20 The complete length ORF91a nucleotide sequence [<SEQ ID 349>] (SEQ ID NO: 349) is:

	1	ATGAAAAAAT	CCTCCTTCAT	CAGCGCATTG	GGCATCGGTA	TTTTGAGCAT
	51	.CGGCATGGCA	TTTGCCGCCC	CTGCCGACGC	GGTAAACCAA	ATCGGTCAAA
	101	ACGCCACTCA	AGTATTGAGC	ATCTTAAAAA	GCGGTGATGC	CAACACCGCC
25	151	CGCCAAAAAG	CCGAAGCCTA	TGCGATTCCG	TATTTGATT	TCCAACGTAT
	201	GACCGCATTG	GCGGTCGGCA	ACCTTGGCG	CACCGGTCC	GACGCGCAAA
	251	AACAAGCGTT	GGCCAAAGAA	TTTCAAACCC	TGCTGATCCG	CACCTATTCC
	301	GGCACGATGC	TGAAATTAAA	AAACGCCAAC	GTCAACGTCA	AAGACAATCC
	351	CATCGTCAAT	AAAGGCGGCA	AAGAAATCAT	CGTCCGCGCC	GAAGTCGGCG
	401	TACCCGGGCA	AAAACCCGTC	AACATGGACT	TCACCACCTA	CCAAAGCGGC
30	451	GGTAAATACC	GTACCTACAA	CGTCGCCATC	GAAGGCGCGA	GCCTGGTTAC
	501	CGTGATCCGC	AACCAATTCG	GCGAAATTAT	CAAGGCGAAA	GGCGTGGACG
	551	GACTGATTGC	CGAGTTGAAG	GCTAAAAACG	GCAGCAAGTA	A

This encodes a protein having amino acid sequence [<SEQ ID 350>] (SEQ ID NO: 350):

35 1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
 51 RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTL LIRTYS
 101 GTMLKLNAN VNVKDNPIVN KGGKEIIVRA EVGVP GQKPV NMDFTTYQSG
 151 GKYRTYNVAI EGASLVTVYR NOFGEI I KAK GVDGLIAELK AKNGSK*

- 40 ORF91a (SEQ ID NO: 350) and ORF91-1 (SEQ ID NO: 348) show 98.0% identity in 196 aa overlap:

orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRONATQVLSILKSGDANTARQKAEAYAIP

[illegible]

Homology with a predicted ORF from *N.gonorrhoeae*

- 20 ORF91 (SEQ ID NO: 346) shows 84.8% identity over a 92aa overlap with a predicted ORF
(ORF91.ng) (SEQ ID NO: 352) from *N. gonorrhoeae*:

	orf91.pep	MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP	60
		: : : : : : : :	
	orf91.ng	VKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP	60
25	orf91.pep	YFDFQRM TALAVGNPWXTXSDXQKQALAXEFQP	93
	orf91.ng	YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLLIRTSYGTMLKFKNATVNVKDNPIVN	120

- The complete length ORF91ng nucleotide sequence [[SEQ ID 351](#)] ([SEQ ID NO: 351](#)) is
30 predicted to encode a protein having amino acid sequence [[SEQ ID 352](#)] ([SEQ ID NO: 352](#)):

1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51 RPKAEAYAVP YFDFQRTAL AVGNPWRTAS DAQQQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEIIKAK GIDGLIAELK AKNGGK*

- Further work revealed the complete nucleotide sequence [**<SEQ ID 353>**] (**SEQ ID NO: 353**):

1	ATGAAAAAAT	CCTCCTTCAT	CAGCGCATTG	GGCATCGGTA	TTTTGAGCAT
51	CGGCATGGCA	TTTGCCCTCC	CGGCCGACGC	AGTGGGACAA	ATCCGCCAAA
101	ACGCCACACA	GGTTTTGACC	ATCCTCAAAA	GCGGCGACGC	GGCTTCTGCA
151	CGCCCAAAAG	CCGAAGCCTA	TGCGGTTCCT	TATTTCGATT	TCCAACGTAT
201	GACCGCATTG	GCGGTCGGCA	ACCCTTGGCG	TACCGCGTCC	GACGCGCAAA
251	ACAACGCGTT	GGCCAAAGAA	TTTCAAACCC	TGCTGATCCG	CACCTATTCC
301	GGCAGCATGC	TGAAATTCAA	AAACGCCGAC	GTCAACGTCA	AAGACAATCC
351	CATCGTCAAT	AAGGGCGGCA	AGGAATACTG	CGTCCGTGCC	GAAGTCGGCA

401 TCCCCGGTCA GAAGCCCGTC AATATGGACT TTACCACCTA CCAAAGCGGC
 451 GGCAAATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC
 501 CGTGATCCGC AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG
 551 GGCTGATTGC CGAGTTGAAA GCCAAAAACG GCGGCAAATA A

This corresponds to the amino acid sequence [<SEQ ID 354; ORF91ng-1>] (SEQ ID NO: 354; ORF91ng-1):

1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
 51 RPKAEAYAVP YPDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTL LIRTY S
 101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
 151 GKYRTYNVAI EGTSLVTYR NQFGEI I KAK GIDGLIAELK AKNGGK*

ORF91ng-1 (SEQ ID NO: 354) and ORF91-1 (SEQ ID NO: 348) show 92.3% identity in 196 aa overlap:

		10	20	30	40	50	60
orf91-1.pep	MKKSSLISALGIGILSIGMAFASPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP						
orf91ng-1	MKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP						
		10	20	30	40	50	60
		70	80	90	100	110	120
orf91-1.pep	YPDFQRM TAL AVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKLNANVNVKDNPIVN						
orf91ng-1	YPDFQRM TAL AVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPIVN						
		70	80	90	100	110	120
		130	140	150	160	170	180
orf91-1.pep	KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYRNQFGEI I KAK						
orf91ng-1	KGGKEIIVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTYRNQFGEI I KAK						
		130	140	150	160	170	180
		190					
orf91-1.pep	GVDGLIAELKAKNGGKX						
orf91ng-1	GIDGLIAELKAKNGGKX						
		190					

In addition, ORF91ng-1 (SEQ ID NO: 354) shows homology to a hypothetical *E.coli* protein (SEQ ID NO: 1136):

sp|P45390|YRBC_ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC REGION
 PRECURSOR (F211))gi|606130 (U18997) ORF_f211 [Escherichia coli])gi|1789583
 (AE000399) hypothetical 24.0 kD protein in murZ-rpoN intergenic region [Escherichia
 coli]Length = 211

Score = 70.6 bits (170), Expect = 6e-12
 Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

Query: 59 VPYPDFQRM TAL AVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPI 118
 +PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
 Sbjct: 65 LPYVQVKYAGALV L GQYYKSATPAQREAYFAAFREYLKQAYGQALAMYHGQTYQIA--PE 122

Query: 119 VNKGKKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQFG 174
 G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
 Sbjct: 123 QPLGDKTIVPIRVTIIDPNGRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

Query: 175 EIIKAKGIDGLIAELKA 191
 +++ KGIDGL A+LK+
 Sbjct: 183 TLLRTKGIDGLTAQLKS 199

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 42

The following DNA sequence was identified in *N.meningitidis* [<SEQ ID 355>] (SEQ ID NO: 355):

```

1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
101 TGATCACGCA TACCCTCATC TCAAAATACA GTTTTGnnn nnnnnnnnnn
151 nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
201 CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGGCAA
251 AAGTCATCGT CTTCCGCACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
301 GACCCCGCCT TCGCCCTGCA ACTGCCCTTA CGCGTCCTCG TTACCGAAAC
351 GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
401 GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
451 AAACGTATAC AAAAAACCGT AGGCGAATAA

```

This corresponds to the amino acid sequence [<SEQ ID 356; ORF97>] (SEQ ID NO: 356; ORF97):

```

1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX
51  XXXXAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
101 DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
151 KLIQKTVGE*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 357>] (SEQ ID NO: 357):

```

1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCCAAAAC GAAACCGCTA
101 TGACCAACGCA TACCCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
201 CGACCATCAG GAAGCCGCCC GCCGAAACGG CTTAACGATG CAGCCGGCAA
251 AAGTCATCGT CTTCCGCACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
301 GACCCCGCCT TCGCCCTGCA ACTGCCCTTA CGCGTCCTCG TTACCGAAAC
351 GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
401 GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
451 AAACGTATAC AAAAAACCGT AGGCGAATAA

```

This corresponds to the amino acid sequence [<SEQ ID 358; ORF97-1>] (SEQ ID NO: 358; ORF97-1):

```

1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTHTLT SKYSFDETVS
51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
101 DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
151 KLIQKTVGE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF97 (SEQ ID NO: 356) shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) (SEQ ID NO: 360) from strain A of *N. meningitidis*:

```

15 orf97.pep      10      20      30      40      50      60
      MKHILPLIAASALCISTASAHHPASEPSTQNETAMITHTLISKYSFGXXXXXXXXXAISKSG
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf97a  MXHILPLXXASALCISTASXHPASEPQTQNETAMTHTLTLSKYSFDETVSRLETAIKSKG
      10      20      30      40      50      60

20 orf97.pep      70      80      90      100     110     120
      MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLTETDGK
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf97a  MDIFAVIDHQAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK
      70      80      90      100     110     120

25 orf97.pep      130     140     150     160
      VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf97a  VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
      130     140     150     160

```

The complete length ORF97a nucleotide sequence [<SEQ ID 359>] (SEQ ID NO: 359) is:

```

30 1  ATGANACACA TACTCCCCCT GANTGNCGCA TCCGCACTCT GCATTTCAAC
    51 CGCTTCGGNN CATCCTGCCA GCGAACCGCA AACCCAAAAC GAAACCGCTA
   101 TGACCACGCA TACCCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC
   151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
   201 CGACCATCAG GAAGCCGCCC GCCGAAACGG CTTAACGATG CAGCCGGCAA
   251 AAGTCATCGT CTTCGGCACG CCCAAAGCCG GTACGCCGCT GATGGTCAAA
   35 GACCCCGCCT TCGCCCTGCA ACTGCCCTG CGCGTCNTCG TTACCGAAAC
   351 GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
   401 GCAGCCGCAT CGGTTTCGAC GAAGTGCGAA ACACTTGGC AAACGCCGAA
   451 AAAGTGATAC AAAAAACCAT AGGCGAATAA

```

This encodes a protein having amino acid sequence [<SEQ ID 360>] (SEQ ID NO: 360):

```

40 1  MXHILPLXXA SALCISTASX HPASEPQTQN ETAMTHTLT SKYSFDETVS
    51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
   101 DPAFALQLPL RVXVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
   151 KLIQKTIGE*

```

ORF97a (SEQ ID NO: 360) and ORF97-1 (SEQ ID NO: 358) show 95.6% identity in 159 aa overlap:

```

5      10      20      30      40      50      60
orf97a.pep  MXHILPLXXASALCISTASXHPASEPQTQNETAMTHTLTSKYSFDETVSRLETAIKSKG
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf97-1      MKHILPLIAASALCISTASAHASEPSTQNETAMTHTLTSKYSFDETVSRLETAIKSKG
           10      20      30      40      50      60

10     70      80      90      100     110     120
orf97a.pep  MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf97-1      MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVETETDGK
           70      80      90      100     110     120

15     130     140     150     160
orf97a.pep  VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf97-1      VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
           130     140     150     160

```

Homology with a predicted ORF from *N.gonorrhoeae*

20 ORF97 (SEQ ID NO: 356) shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) (SEQ ID NO: 362) from *N. gonorrhoeae*:

```

orf97.pep  MKHILPLIAASALCISTASAHASEPSTQNETAMITHLISKYSFGXXXXXXXXXAISKSG 60
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf97ng     MKHILPPIAASAFICISTASAHAGKPPTQNETAMTHTLTSKYSFDETVSRLETAIKSKG 60

25  orf97.pep  MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVETETDGK 120
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf97ng     MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVETETDGK 120

orf97.pep  VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGE 159
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf97ng     VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGE 159
30

```

The complete length ORF97ng nucleotide sequence [<SEQ ID 361>] (SEQ ID NO: 361) is predicted to encode a protein having amino acid sequence [<SEQ ID 362>] (SEQ ID NO: 362):

```

35      1  MKHILPPIAA SAFCISTASA HPAGKPPTQN ETAMTHTLT SKYSFDETVS
      51  RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
     151  KLIQKTVGE*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 363>] (SEQ ID NO: 363):

```

40      1  ATGAAACACA TACTCCCcct gatcgccgca TccgcactCT GCATTTC AAC
     51  CGCTTCGGCA CACCCTGCCG GCAAACCGCC CACCCAAAAC GAAACCGCTA

```

101 TGACCACGCA CACCCTCACC TCGAAATACA GTTTTGACGA AACCGTCAGC
 151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT
 201 CGACCATCAG GAAGCGGCAC GCCGAAACGG CCTGACCATG CAGCCGGCAA
 251 AAGTCATCGT CTTCCGGCACG CCCAAGGCCG GTACGCCGct GATGGTCAAA
 301 GACCCCGCCT TCGCCCTGCA ACTGCCCTTG CGCGTCCTCG TTACCGAAAC
 351 GGACGGCAA GTACGCACCG CCTATACCGA TACGCGCGCC CTCATCGTCG
 401 GCAGCCGCAT CAGTTTCGAC GAAGTGGCAA AACTTTGGC AAACGCCGAA
 451 AAAGTGATAC AAAAAACCGT AGGCGAATAA

10 This corresponds to the amino acid sequence [<SEQ ID 364; ORF97ng-1>] (SEQ ID NO: 364; ORF97ng-1):

1 MKHILPLIAA SALCISTASA HPAGKPPTQN ETAMTTHTLT SKYSFDETVS
 51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
 101 DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
 151 KLIQKTVGE*

ORF97ng-1 (SEQ ID NO: 364) and ORF97-1 (SEQ ID NO: 358) show 96.2% identity in 159 aa overlap:

20	orf97-1.pep	10	20	30	40	50	60
		MKHILPLIAAASALCISTASAHPASEPSTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG					
	orf97ng-1	MKHILPLIAAASALCISTASAHPAAGKPPTQNETAMTTHTLTSKYSFDETVSRLETAIKSKG					
		10	20	30	40	50	60
25	orf97-1.pep	70	80	90	100	110	120
		MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK					
	orf97ng-1	MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK					
		70	80	90	100	110	120
30	orf97-1.pep	130	140	150	160		
		VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX					
		: : :					
	orf97ng-1	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX					
		130	140	150	160		

35 Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (SEQ ID NO: 358) (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by
 40 SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure

12D). These experiments confirm that ORF97-1 (SEQ ID NO: 358) is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1 (SEQ ID NO: 358).

5 Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* [SEQ ID 365>] (SEQ ID NO: 365):

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC
51  GCTGATGCTC CCCGCTTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA
101 GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
151 CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGg
201 CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCCGCC CCGATAATCG
251 CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACaATATT
301 GACTACAAAC TGAGTTTCCA TCCGCTGACc AaACGCTACC GCGTTACCgT
15  351 CGgCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
401 CCGGCGCGGT TGCCAACCTGG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
451 GCGGAAGCAG GGGAAACCAA GCGGAAATC CGCCTGACGC TGTCCACTTC
501 AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
20  551 ATTTGGATTc GGGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA

```

This corresponds to the amino acid sequence [SEQ ID 366; ORF106>] (SEQ ID NO: 366; ORF106):

```

1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
51  RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYRFLK GQLIGDDNDI
25  101 DYKLSFHPLT KRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
151 AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWHLDSGWK PLNIIGNK*

```

Further work revealed the following DNA sequence [SEQ ID 367>] (SEQ ID NO: 367):

```

30  1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC
51  GCTGATGCTC CCCGCTTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA
101 GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
151 CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TGCGCCGGGG
201 CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCCGCC CCGATAATCG
251 CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACAATATT
35  301 GACTACAAAC TGAGTTTCCA TCCGCTGACC AACCCTTACC GCGTTACCGT
351 CGGCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
401 CCGGCGCGGT TGCCAACCTGG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
451 GCGGAAGCAG GGGAAACCAA GCGGAAATC CGCCTGACGC TGTCCACTTC
501 AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACCTGGC
40  551 ATTTGGATTc GGGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA

```

This corresponds to the amino acid sequence [<SEQ ID 368; ORF106-1>] (SEQ ID NO: 368; ORF106-1):

```

5      1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
      51  RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYRFLK GQLIGDDDDNI
     101  DYKLSFHPLT NRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
     151  AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWHLD SGWK PLNIIGNK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF106 (SEQ ID NO: 366) shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) (SEQ ID NO: 370) from strain A of *N. meningitidis*:

```

15      orf106.pep      10      20      30      40      50      59
      MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ
      ||||| ||| |:: || :: ||||| ||||| : ||||| ||||| |||||
      orf106a          MAFITRLFKSIKQWLVLPLMLSVLPDAAAEGIDVSRAEARIXDGGQLSXXSRFQTELPDQ
      10      20      30      40      50      60

      orf106.pep      60      70      80      90      100      110      119
      LQQALRRGVPLNFTLSWQLSAPIIASYRFLKLGQLIGDDDNIDYKLSFHPLTKRYRVTVGA
      || | ||| || | ||||| ||||| ||||| ||||| ||||| ||||| : ||||| |||||
      20      orf106a      LQXAXXRGVXLNXTLXWQLSAPIIASYRFXLGLGDDDXIDYKLSFHPLTNRYRVTVGA
      70      80      90      100      110      120

      orf106.pep      120      130      140      150      160      170      179
      FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      25      orf106a      FSTXYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT
      130      140      150      160      170      180

      orf106.pep      180      190      199
      SQNWHLD SGWKPLNIIGNKX
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      30      orf106a      SQNWHLD SGWKPLNIIGNKX
      190      200

```

Due to the K→N substitution at residue 111, the homology between ORF106a (SEQ ID NO: 370) and ORF106-1 (SEQ ID NO: 368) is 87.9% over the same 199 aa overlap.

35 The complete length ORF106a nucleotide sequence [<SEQ ID 369>] (SEQ ID NO: 369) is:

```

40      1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
      51  GCTGCCGATG CTTTCCGTTT TGCCGGACGC GCGGCGGAG GGGATAGATG
     101  TGAGCCGCGC CGAAGCGAGG ATAANCGACG GCGGGCAGCT TTCCATNAGN
     151  AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAANNNG CGNNGNGCCG
     201  GGGCGTGNCG CTCAACTNTA CCTTAAGNTG GCAGCTTTCC GCCCCGATAA
     251  TCGCTTCTTA TCGGTTTNA TGGGGCAAC TGATTGGCGA TGACGACNAT

```

5

301	ATTGACTACA	AACTGAGTTT	CCATCCGCTG	ACCAACCGCT	ACCGCGTTAC
351	CGTCGGCGCG	TTTTCGACAG	ANTACGACAC	CTTGGATGCG	GCATTGCGCG
401	CGACCGGCGC	GGTGCGCAAC	TGGAAAGTCC	TGAACAAAGG	CGCGCTGTCC
451	GGTGCGGAAG	CAGGGGAAAC	CAAGGCGGAA	ATCCGCTGTA	CGCTGTCCAC
501	TTCAAAATCG	CCCAAGCCTT	TTCAAATCAA	TGCATTGACT	TCTCAAAACT
551	GGCATTTTGGA	TTCCGGGTTG	AAACCTCTAA	ACATCATCGG	GAACAAATAA

This encodes a protein having amino acid sequence [<SEQ ID 370>] (SEQ ID NO: 370):

10

1	MAFITRLFKS	IKQWLVL LPM	LSVLPDAAAE	GIDVSRAEAR	IXDGGQLSXX
51	SRFQTELPDQ	LQXAXXRGVX	LNXTLXWQLS	APIIASYRFX	LGQLIGDDDX
101	IDYKLSFHPL	TNRYRVTVGA	FSTXYDTLDA	ALRATGAVAN	WKVLNKGALS
151	GAEAGETKAE	IRLTLSTSKL	PKPFOINALT	SONWHLDSGW	KPLNIIGNK*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF106 (SEQ ID NO: 366) shows 90.5% identity over a 199aa overlap with a predicted ORF
15 (ORF106.ng) (SEQ ID NO: 372) from *N. gonorrhoeae*:

	orf106.pep	MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLS ISSRFQTELPDQ	59
	orf106ng	MAFITRLFKSIKQWLVLPLILSVLPDAAAEGIAATRAEARITDGGRLS ISSRFQTELPDQ	60
20	orf106.pep	LQQALRRGVPLNFTLSWQLSAPIIASYRFLKGLIGDDDNIDYKLSFHPLTKRYRVTVGA	119
	orf106ng	LQQALRRGVPLNFTLSWQLSAPTIIASYRFLKGLIGDDDNIDYKLSFHPLTNRYRVTVGA	120
	orf106.pep	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT	179
	orf106ng	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT	180
25	orf106.pep	SQNWHLD SGWKPLNIIGNK	198
	orf106ng	SONWHLD SGWKPLNIIGNK	199

Due to the K→N substitution at residue 111, the homology between ORF106ng (SEQ ID NO: 372) and ORF106-1 (SEQ ID NO: 368) is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence [<SEQ ID 371>] (SEQ ID NO: 371) is:

	1	ATGGCTTTTA	TTACGCGCTT	ATTCAAAGC	ATTAACAAT	GGCTTGTGCT
	51	GTTGCCGATA	CTCTCCGTTT	TGCCGGACGC	GGCGGCGGAG	GGCATTGCCG
35	101	CGACCCGCGC	CGAAGCGAGG	ATAACCGACG	GCGGGCGGCT	TTCCATCAGC
	151	AGCCGCTTCC	AAACCGAGCT	GCCCAGACCAG	CTCCAACAGG	CGTTGCGCCG
	201	GGCGTACC	CTCAACTTTA	CCTTAAGCTG	GCAGCTTTCC	GCCCCGACAA
	251	TCGTTCTTTA	TCGGTTTAAA	TTGGGGCAAC	TGATTGGCGA	TCCAGACAAT
	301	ATTGACTACA	AACTAAGTTT	CCATCCGCTG	ACCAACCGCT	ACCGCGTTAC
	351	CGTCGGCGCA	TTTTCCACCG	ATTACGACAC	TTTGGATGCG	GCATTGCGCG
40	401	CGACCGGCGC	GGTTGCCAAC	TGGAAAGTCC	TGAACAAAGG	CGCGTTGTCC
	451	GGTGCGGAAG	CAGGGGAAAC	CAAGGCGGAA	ATCCGCCTGA	CGCTGTCCAC
	501	TTCAAACTG	CCCAAGCCTT	TCCAAATCAA	CGCATTGACT	TCTCAAAACT
	551	GGACTTGGCA	TTCCGGTTGG	AAACCTCTAA	ACATCATCGG	GAACAAATAA

This encodes a protein having amino acid sequence [<SEQ ID 372>] (SEQ ID NO: 372):

```

5      1  MAFITRLFKS IKQWLVLPI LSVLPDAAAE GIAATRAEAR ITDGGRLSIS
      51  SRFQTELPDQ LQQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
      101 IDYKLSFHPL TNRYRVTVGA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
      151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLD SGW KPLNIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (SEQ ID NO: 368) (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 (SEQ ID NO: 368) is a surface-exposed protein, and that it is a useful immunogen.

Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 373>] (SEQ ID NO: 373):

```

20      1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
      51  GGT'TTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG
      101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG GCTgACGGTG
      151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
25      201 CACCGCCGAC AAAGACAcCT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
      251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
      301 TCTGAAATCC TGTTTTCAC TCGACGATGCC gCCGCCGGCa TCGGGCTGGT
      351 GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTGC
      401 GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCCcAAG
30      451 CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
      501 AGCGAACACC GCCGTCTGA CCGCGCTTTA CGCGCTGGCA AACCTTGCCG
      551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
      601 CACGCACCGT TTTGCCCGC CGTCCTGCAC CGGGGG.TGC GCTACGGCAT
      651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
35      701 GTTTGTTCCT GAAAAAATAT GCCGCGCTGG AACAGCTCGG CGTTTATTCTG
      751 ATGGGTATTT CGTTCGGCGG GCGGCGATTA TTGTTCAAA GCATCTTTTC
      801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
      851 CCGCTCGCCT CTCGGCAACG GCAGAATCCG CCGCGCCCT GCTTGCCCTC
      901 GCCCTCTGC.TGACCGGCAT TTTCTCGCCC CTGCTCTCCC TCCTGCTGCC
40      951 GGAAACTAC GCCCGCTCC GGTATATCGT CGTATCGTGT ATG.TGCCGC
      1001 CGCTGTTTTG CACGTGGCG GAAATCAGCG GCATCGGTTT GAACGTGCTT

```

5
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGGCGAGG CCGCC.GGCG
1151 CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTTTG CTTCAAGACC
1201 GAAAGCTCyT GCCGCCTGTG GCAGCCGCTC AAACGCCTGC CGCTTTATCT
1251 GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
1301 CGCCGCAAAA CTATCCCCTG TTTGCCGGCG TATGGGCGGC ATATCTGGCA
1351 GGCTGCATCC TGCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT
1401 GAAAAACAA GGTTCCTCAT TATGA

10 This corresponds to the amino acid sequence [<SEQ ID 374; ORF10>] (SEQ ID NO: 374;
ORF10):

15
20
40
45
1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLF E LSFLPIRFL LVLRMGRAL AFSSAQLVPK
151 LAILLXPLT VGLLHFPANT AVLTAVYALA NLAAAAFLLF QNRCRLKAVR
201 HAPFPAVLH RGXYGIPIA LSSIAWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCXTGIFSP LASLLLPENY AAVRFIVVSC MXPPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLDRAVPAR PXGAAVACAA SFWLFFAFKT
401 ESSCRLWQPL KRLPLYLHTL FCLTSSAAYT CFGTPANYPL FAGVWAAyla
451 GCILHRKDL HKLFHYLKKQ GFPL*

Further sequence analysis revealed the complete DNA sequence [<SEQ ID 375>] (SEQ ID NO:
375) to be:

25
30
35
40
45
50
1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCGCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CCGCGCGGG GCTGACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAAGACACCT TGTTCAAAC CCTGTTCTTG CCGCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTCACT CGACGATGCC GCCGCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTC
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTTC
501 AGCGAACACC GCCGCTCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCCT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTCCGCCGC CGTCTGCAC CCGGGGCTGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCTT GAAAAATAT GCCGCGCTGG AACAGCTCGG CGTTTATTCG
751 ATGGGTATTT CGTTCGGCGG GCGGCGATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCCCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCCTC
901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTTGCCTCCC TCCTGCTGCC
951 GGAAACTAC GCCGCCGTCC GGTTTATCGT CGTATCGTGT ATGCTGCCGC
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGCAAAACGC GCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTGCCGTC CCGCGGCGCG CGCGGCGCGG
1151 CGGTTGCCTG TGCCGCCTCA TTCTGGCTGT TTTTTCCTT CAAGACCGAA
1201 AGCTCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCGC TTTATCTGCA
1251 CACATTGTTT TGCCGTGACCT CCTCGGCGGC CTACACCTGC TTCGGCACGC
1301 CGGCAAACTA TCCCCTGTTT GCCGGCGTAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCTGTC GCCACCGGAA AGATTTCAC AACTGTTC ATTATTTGAA
1401 AAAACAAGGT TTCCATTAT GA

This corresponds to the amino acid sequence [<SEQ ID 376; ORF10-1>] (SEQ ID NO: 376; ORF10-1):

```

      1  MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
      51  SVLCLGLDQA YVREYYATAD KDTLTKTLFL PLLSAAAIA ALLLSRPSLP
5      101  SEILFSLDDA AAGIGLVLF LSF LPIRFL LVL RMEGRAL AFSSAQLVPK
      151  LAII LLLPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
      201  HAPFSPAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
      251  MGISFGGAAL LFQSIFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
      301  ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPPLFCTLA EISGIGLNVV
10     351  RKTRPIALAT LGALANLLL LGLAVPSGGA RGA AVACAAS FWLFFAFKTE
      401  SSCRLWQPLK RLPLYLHTLF CLTSSAAYTC FGTPANYPLF AGVWAAYLAG
      451  CILRHRKDLH KLFHYLKKQG FPL*
```

Computer analysis of this amino acid sequence gave the following results:

15 Prediction

ORF10-1 (SEQ ID NO: 376) is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

20 Homology with EpsM (SEQ ID NO: 1137) from *Streptococcus thermophilus* (accession number U40830).

ORF10 (SEQ ID NO: 374) shows homology with the epsM gene of *S. thermophilus*, which encodes a protein (SEQ ID NO: 1137) of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

```

25  Query:   213  LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVYSMGISFGGAALLLQSIFSTVW 270
      L Y +PL SS+ +W L ++ R F+ + G   G+ ++           + +IF+ W
      Sbjct:  210  LYYALPLIPSSILWLLNASSRYFVLFFLGAGANGLLAVATKIPSIISIFNTIFTQAW 267
```

Identities = 15/57 (26%), Positives = 31/57 (54%)

```

30  Query:    7  LGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63
      L +  G++GS +L  +++PL ++      + G   L QT A L + ++ + +  A +R
      Sbjct:   12  LVFTIGNLGSKLLVFLLVPLYTYAMTPQEYGMADLYQTTANLLLPLITMNVFDTATLR 68
```

Identities = 16/96 (16%), Positives = 36/96 (37%)

```

35  Query:  307  IFSPLASLLLPENYA AVRFTVVSCLMPPLFYTLTEISGIGLNVVRKTRPIXXXXXXXXXX 366
      + P+  ++ +YA+   V  ML  LF + ++ G      ++T+ +
      Sbjct:  305  VLKPIVEKVVSSDYASSWQYVPFFMLSMLFSSFSDFGTNYIAAKQTKGVFMTSIYGTIV 364
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF10 (SEQ ID NO: 374) shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) (SEQ ID NO: 378) from strain A of *N. meningitidis*:

5	orf10.pep	10 20 30 40 50 60	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
	orf10a	10 20 30 40 50 60	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
10	orf10.pep	70 80 90 100 110 120	YVREYYATADKDTLTKTLFLPPLLSAAAIALLSRPSLPSEILFSLDDAAAGIGLVLFEE
	orf10a	70 80 90 100 110 120	YVREYYAAADKDTLTKTLFLPPLLSAAAIALLSRPSLPSEILFSLDDAAAGIGLVLFEE
15	orf10.pep	130 140 150 160 170 180	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA
	orf10a	130 140 150 160 170 180	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLPLTVGLLHFPANTAVLTAVYALA
20	orf10.pep	190 200 210 220 230 240	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY
	orf10a	190 200 210 220 230 240	NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAYWGLASADRLFLKKY
25	orf10.pep	250 260 270 280 290 300	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS
	orf10a	250 260 270 280 290 300	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS
30	orf10.pep	310 320 330 340 350 360	ALCXTGIFSPLASLLLPENYAAVRFIVVSCMXPPLFCTLAIEISGIGLNVVRKTRPIALAT
	orf10a	310 320 330 340 350 360	ALCLTGIFSPLASLLLPENYAAVRFIVVSCMLPPLFCTLVEISGIGLNVVRKTRPIALAT
35	orf10.pep	370 380 390 400 410 419	LGALAAANLLLLGLDRAVPAR-PXGAAVACAASFVWFFFAFKTESSCRLWQPLKRLPLYLHT
	orf10a	370 380 390 400 410	LGALAAANLLLLGL--AVPSGGARGAAVACAASFVWFFVFKTESSCRLWQPLKRLPLYMHT
40	orf10.pep	420 430 440 450 460 470	LFCLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLPHYLLKKQGFPLX
	orf10a	420 430 440 450 460 470	LFCLASSAAYTCFGTPANYPLFAGVWAVYLAGCILRHRKDLHKLPHYLLKKQGFPLX

45 The complete length ORF10a nucleotide sequence [<SEQ ID 377>] (SEQ ID NO: 377) is:

```

      1  ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
     51  GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCTGCCG
    101  ACGACATCGG ACGCATCGTG CTGATGCAGA CGGCGGCGGG GCTGACGGTG
    151  TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
    201  CGCCGCCGAC AAAGACACTT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
    251  TGTCTGCCGC CGCATAGACC GCCCTGCTGC TTTCCCGCCC ATCCCTGCCG
    301  TCTGAAATCC TGTTTTCGCT CGACGATGCC GCCGCCGGCA TCGGGCTGGT
    351  GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG
    401  GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCAAAG
    451  CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
    501  GGCGAACACC GCCGTCCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
    551  CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGCTGAA GGCCGTCCGG
    601  CGCGCACCGT TTTATCCGC CGTCCTGCAT CGCGGCCTGC GTACGGCAT
    651  ACCGATCGCA CTAAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
    701  GTTTGTTCCT GAAAAAATAT GCCGGCCTAG AACAGCTCGG CGTTTATTCG
    751  ATGGGTATTT CGTTCGGCGG AGCGGCATTA TTGTTCCAAA GCATCTTTTC
    801  AACGCTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGCA AACGCCCCGC
    851  CCGCCCGCCT CTCGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC
    901  GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTCGCCTCCC TCTGCTGCC
    951  GGAAACTAC GCCGCGTCC GGTATATCGT CGTATCGTGT ATGCTGCCTC
   1001  CGCTGTTTTG CACGCTGGTA GAAATCAGCG GCATCGGTTT GAACGTCGTC
   1051  CGAAAAACAC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
   1101  CCTGCTGCTG CTGGGGCTTG CCGTACCGTC CGGCGGCGCG CGCGGCGCGG
   1151  CGGTTGCCTG TGCCGCTCA TTTTGCTGT TTTTGTGTTT CAAGACCGAA
   1201  AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCGC TTTATATGCA
   1251  CACATTGTTC TGCTTGGCCT CCTCGGCGGC CTACACCTGC TTCGGCACTC
   1301  CGGCAAACCTA CCCCCTGTTT GCCGCGGTAT GGGCGGTATA TCTGGCAGGC
   1351  TGCATCCTGC GCCACCGGAA AGATTGTCAC AAAGTGTTC ATTATTTGAA
   1401  AAAACAAGGT TTCCATTAT GA

```

This encodes a protein having amino acid sequence [SEQ ID 378] (SEQ ID NO: 378):

```

      1  MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
     51  SVLCLGLDQA YVREYYAAAD KDTLFKTLFL PPLLSAAAIA ALLLSRPSLP
    101  SEILFSLDDA AAGIGLVLF LSFPIRFL LVLMEGRAL AFSSAQLVSK
    151  LAILLLLPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
    201  RAPFSSAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
    251  MGISFGGAAL LFQSIPTVW TPYIFRAIEA NAPPARLSAT AESAAALLAS
    301  ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPPLFCTLV EISGIGLNVV
    351  RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFVFKTE
    401  SSCRLWQPLK RLPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAVYLAG
    451  CILRHRKDLH KLFHYLKKQG FPL*

```

ORF10a (SEQ ID NO: 378) and ORF10-1 (SEQ ID NO: 376) show 95.4% identity in 475 aa overlap:

```

    45      10      20      30      40      50      60
    orf10-1.pep  MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
    orf10a       MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
                  10      20      30      40      50      60

    50      70      80      90     100     110     120
    orf10-1.pep  YVREYYATADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE
    orf10a       YVREYYAAADKDTLFKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE
                  70      80      90     100     110     120

```

-310-

		130	140	150	160	170	180
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA					
5	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLXPLTVGLLHFPANTAVLTAVYALA					
		130	140	150	160	170	180
	orf10-1.pep	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY					
10	orf10a	NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAYWGLASADRLFLKKY					
		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
15	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					
		250	260	270	280	290	300
	orf10-1.pep	ALCXTGIFSPLASLLLPENYAARFIVVSCMXPPFLCTLAIEISGIGLNVVRKTRPIALAT					
20	orf10a	ALCLTGIFSPLASLLLPENYAARFIVVSCMLPPLFCTLVEISGIGLNVVRKTRPIALAT					
		310	320	330	340	350	360
	orf10-1.pep	LGALAANLLLLGLDRAVPAR-PXGAAVACAASFVWFFAFKTESSCRLWQPLKRLPLYLHT					
25	orf10a	LGALAANLLLLGL--AVPSGGARGAAVACAASFVWFFVFKTESSCRLWQPLKRLPLYMHT					
		370	380	390	400	410	419
	orf10-1.pep	LFCLTSSAAYTCFGTPANYPLFAGVWAAVLGACILRHRKDLHKLPHYLLKKQGFPLX					
30	orf10a	LFCLASSAAYTCFGTPANYPLFAGVWAVVLGACILRHRKDLHKLPHYLLKKQGFPLX					
		420	430	440	450	460	470

Homology with a predicted ORF from *N.gonorrhoeae*

ORF10 (SEQ ID NO: 374) shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) (SEQ ID NO: 380) from *N. gonorrhoeae*:

35	orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
	orf10ng.pep	YVREYYAAADKDTLTKTLFLPPLLFSAIIAALLSRPSLPSEILFSLDDAAAGIGLVLF	120
	orf10nm	YVREYYATADKDTLTKTLFLPPLLSAAIIAALLSRPSLPSEILFSLDDAAAGIGLVLF	120
40	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTSVLTAUYALA	180
	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA	180
	orf10ng.pep	NLAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPLALSSLAYWGLASADRLFLKKY	240
45	orf10nm	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY	240

orf10ng.pep AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS 300
 orf10nm AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS 300
 5 orf10ng.pep ALCLTGIFSPLASLLLPENYA AVRFTVVS CMLPPLFYTLTEISGIGLN VVRKTRPIALAT 360
 orf10nm ALCLTGIFSPLASLLLPENYA AVRFTVVS CMLPPLFYTLTEISGIGLN VVRKTRPIALAT 360
 10 orf10ng.pep 370 380 390 400 410
 LGALAANLLLLGL--AVPSGGTRGA AVACAASFWLFFVFKTESSCRLWQPLKRLPLYMHT
 orf10nm LGALAANLLLLGLDRAVPAR-PXGA AVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT
 370 380 390 400 410
 15 orf10ng.pep 420 430 440 450 460 470
 LFCLASSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKNLHKLFHYLKKQGFPLX
 orf10nm LFCLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLFHYLKKQGFPLX
 420 430 440 450 460 470

The complete length ORF10ng nucleotide sequence [<SEQ ID 379>] (SEQ ID NO: 379) is:

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
 20 51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCcccgCCG
 101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG ACTGACGGTG
 151 TCGGTATTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
 201 CGCCGCCGAC AAAGACACTT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
 25 251 TGTTTTCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
 301 TCTGAAATCC TGTTTTCGCT CGACGATGCC GCCGCGGCA TCGGGCTGGT
 351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTGC
 401 GTATGGAAGG GCGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAA
 451 CTCGCCATTC TGCTGCTGTT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
 501 GCGGAACACC TCCGTCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
 30 551 CCGCCGCCCT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCG,
 601 CGCGCGCCGT TTTGCCCGC CGTCCTGCAC CGGGGCTGC GCTACGGCAT
 651 ACCGCTCGCA CTGAGCAGCC TTGCCTATTG GGGGCTGGCA TCCGCCGACC
 701 GTTTGTTCCT GAAAAAATAT GCGGGCCTGG AACAGCTCGG CGTTATTTCG
 751 ATGGGTATT CGTTCGGCGG GCGGCATTA TTGCTCCAA GCATCTTTTC
 35 801 AACGGTCTGG ACACGTAATA TTTTCCGTGC AATCGAAGAA AACGCCACGC
 851 CCGCCCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTCC
 901 GCCCTCTGCC TGACCGGAAT TTTCTCGCCC CTCGCCTCCC TCCTGCTGCC
 951 GGAAACTAC GCCGCCGTCC GGTTTACCGT CGTATCGTGT ATGCTGccgc
 1001 cgcTGTTTTA CACGCTGACC GAAATCAGCG GCATCGGTTT GAACGTCGTC
 40 1051 CGAAAACGC GTCCGATCGC GCTTGCCACC TTGGGCGCGC TGGCGGCAAA
 1101 CCTGCTGCTG CTGGGGCTTG CCGTACCGTC CGGCGGCACG CGCGGCGCGG
 1151 CGGTTGCCCTG TGCCGCCTCA TTCTGGTTGT TTTTGTGTTT CAAGACAGAA
 1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAAA CGCCTGCCGC TTTATATGCA
 1251 CACATTGTTC TGCTTgGCCT CCTCGGCGGC CTACACCTGC TTCGCACAC
 45 1301 CGGCAACTA CCCcctgtt gccggcgtAT GGGCGGCATA TCTGGCAGGC
 1351 TGCATCCTGC GCCACCGGAA AAATTTGCAC AAAGTGTTC ATTATTTGAA
 1401 AAAACAAGGT TTCCATTAT GA

This encodes a protein having amino acid sequence [<SEQ ID 380>] (SEQ ID NO: 380):

50 1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
 51 SVLCLGLDQA YVREYYAAD KDTLFKTLFL PLLFSAAIA ALLLSRPSLP
 101 SEILFSLDDA AAGIGLVLE LSFLPIRFL LVLRMGRAL AFSSAQLVPK
 151 LAILLLLPLT VGLLHFPANT SVLTAVYALA NLAAAFLLF QNRCRLKAVR

201 RAPFSPAVLH RGLRYGIPLA LSSLAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LLQSIFSTVW TPYIFRAIEE NATPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFTVWSC MLPPLFYTLT EISGIGLNVV
351 RKTRPIALAT LGALAAANLLL LGLAVPSGGT RGA AVACAAS FWLFFVFKTE
401 SSCRLWQPLK RLPPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAAAYLAG
451 CILRHRKNLH KLFHYLKKQG FPL*

ORF10ng (SEQ ID NO: 380) and ORF10-1 (SEQ ID NO: 376) show 96.4% identity in 473 aa overlap:

10		10	20	30	40	50	60
	orf10-1.pep	MDTKEILGYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLT	TVSVLCLGLDQA			
	orf10ng-1	MDTKEILGYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLT	TVSVLCLGLDQA			
		10	20	30	40	50	60
15		70	80	90	100	110	120
	orf10-1.pep	YVREYYATADKDTLFTKTLFLP	PLLSAAAIAALLSRPSLPSEIL	FSLDDAAAGIGLVLFE			
	orf10ng-1	YVREYYAAADKDTLFTKTLFLP	PLLSAAAIAALLSRPSLPSEIL	FSLDDAAAGIGLVLFE			
		70	80	90	100	110	120
20		130	140	150	160	170	180
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAF	SSAQLVPKLAILLPLTVGLLH	FPANTAVLTAVYALA			
	orf10ng-1	LSFLPIRFLLLVLRMEGRALAF	SSAQLVPKLAILLPLTVGLLH	FPANTSVLTAVYALA			
		130	140	150	160	170	180
25		190	200	210	220	230	240
	orf10-1.pep	NLAAAFLLFQNRCLKAVRHAP	FSPAVLHRLRYGIPIALSS	IAYWGLASADRLFLKKY			
	orf10ng-1	NLAAAFLLFQNRCLKAVRRAP	FSPAVLHRLRYGIPLALSS	LAYWGLASADRLFLKKY			
		190	200	210	220	230	240
30		250	260	270	280	290	300
	orf10-1.pep	AGLEQLGVYSMGISFGGAALL	FQSIFSTVWTPYIFRAIEEN	APPARLSATAESAAALLAS			
	orf10ng-1	AGLEQLGVYSMGISFGGAALL	QSIFSTVWTPYIFRAIEEN	ATPARLSATAESAAALLAS			
		250	260	270	280	290	300
35		310	320	330	340	350	360
	orf10-1.pep	ALCLTGIFSPLASLLLPENY	AAVRFIVVSCMLPPLFCT	LAEISGIGLNVVRKTRPIALAT			
	orf10ng-1	ALCLTGIFSPLASLLLPENY	AAVRFTVVSCMLPPLFYTL	TEISGIGLNVVRKTRPIALAT			
		310	320	330	340	350	360
40		370	380	390	400	410	420
	orf10-1.pep	LGALAAANLLLGLAVPSGG	GARGAAVACAASFWLFF	FAFKTESSCRLWQPLKRL	PLYLHTLF		
	orf10ng-1	LGALAAANLLLGLAVPSGG	TRGA AVACAASFWLFF	VFKTESSCRLWQPLKRL	PLYMHTLF		
		370	380	390	400	410	420
45		430	440	450	460	470	
	orf10-1.pep	CLTSSAAYTCFGTPANYPL	FAGVWAAAYLAGCILRHR	KNLHKLHKLHKLHKLHKLH	KLHKLHKLHKLHKLHKLH		
	orf10ng-1	CLASSAAYTCFGTPANYPL	FAGVWAAAYLAGCILRHR	KNLHKLHKLHKLHKLHKLH	KLHKLHKLHKLHKLHKLH		
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 45

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 381>] (SEQ ID NO: 381):

```

1..ATCCTGAAAC CGCATAACCA GCTTAAGGAA GACATCCAAC CTGATCCGGC
51  CGATCAAAAC GCCTTGTCGG AACCGGATGC TGCACAGAG GCAGAGCAGT
10  101 CCGATGCGGA AAATGCTGCC GACAAGCAGC CCGTTGCCGA TAAAGCCGAC
151 151 GAGGTTGAAG AAAAGGCGGG CGAGCCGGAA CGGGAAGAGC CGGACGGACA
201 201 GGCAGTGCGT AAGAAAGCGC TGACGGAAGA GCGTGAACAA ACCGTCAGGG
251 251 AAAAAGCGCA GAAGAAAGAT GCCGAAACGG TTAAATACA AGCGGTAAAA
301 301 CCGTCTAAAG AAACAGAGAA AAAAGCTTCA AAAGAAGAGA AAAAGGCGGC
15 351 GAAGGAAAAA GTTGACCCCA AACCAACCCC GGAACAAATC CTCAACAGCG
401 401 GCAGCATCGA AAAGGCGCGC AGTGCCGCCG CCAAGAAGT GCAGAAAATG
451 451 AA.AACGTCC GACAAGGCGG AAGC.AACGC ATTATCTGCA AATGGGCGCG
501 501 TATGCCGACC GTCAGAGCGC GGAAGGCGAG CGTGCCAAAC TGGCAATCTT
551 551 GGGCATATCT TCCAAGGTGG TCGGTTATCA GCGGGACAT AAAACGCTTT
20 601 ACCGGGTGCA AAGCGCAAT ATGTCTGCCG ATGCGGTGA

```

This corresponds to the amino acid sequence [<SEQ ID 382; ORF65>] (SEQ ID NO: 382; ORF65):

```

1..ILKPHNQLKE DIQDPADQN ALSEPDAATE AEQSDAENAA DKQPVADKAD
25 51 EVEEKAGEPE REEPDQAVR KKALTEEREQ TVREKAQKKD AETVKIQAVK
101 101 PSKETEEKAS KEEKKAAKEK VAPKPTPEQI LNSGSIEXAR SAAAKEVQKM
151 151 XNVRQGSXR IICKWARMPT VRARKGSVPN WQSWAYLPRW SVIRRDIKRF
201 201 TGCKAAICLP MR*

```

30 Further work revealed the complete nucleotide sequence [<SEQ ID 383>] (SEQ ID NO: 383):

```

1  ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTTTT
51  CTTCCGTTTG ATACTGGCGA CGGTCATTAT TGCCGGTATT TTGTTTATC
101 TGAACCAGAG CCGTCAAAAT GCGTTCAAAA TCCCGGCTTC GTCGAAGCAG
151 CCTGCAGAAA CGGAAATCCT GAAACCGAAA AACCAGCCTA AGGAAGACAT
35 201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGCTGCGA
251 CAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT
301 GCCGATAAAG CCGACGAGGT TGAAGAAAAG GCGGGCGAGC CGGAACGGGA
351 AGAGCCGGAC GGACAGGCAG TGCCTAAGAA AGCGCTGACG GAAGAGCGTG
401 AACAAACCGT CAGGGAAAAA GCGCAGAAGA AAGATGCCGA AACGGTTAAA
40 451 AAACAAGCGG TAAACCGTC TAAAGAAACA GAGAAAAAAG CTTCAAAAGA
501 AGAGAAAAAG GCGGCGAAGG AAAAAGTTGC ACCCAAACCA ACCCGGAAC
551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCCGCCAAA
601 GAAGTGCAGA AAATGAAAAC GTCCGACAAG GCGGAAGCAA CGCATTATCT
651 GCAAATGGGC GCGTATGCCG ACCGTCAGAG CGCGGAAGGG CAGCGTGCCA
45 701 AACTGGCAAT CTTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGGCGGGA

```

751 CATAAAACGC TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT
801 GAAAAAAATG CAGGACGAGT TGAAAAACA TGAAGTCGCC AGCCTGATCC
851 GTTCTATCGA AAGCAAATAA

- 5 This corresponds to the amino acid sequence [SEQ ID 384; ORF65-1] (SEQ ID NO: 384; ORF65-1):

10

1	MFMNKFQSQG	KGLSGFFFGL	ILATVIIAGI	LFYLNQSQGN	AFKIPASSKQ
51	PAETELKPK	NQPKEDIQPE	PADQNALSEP	DAATEAEQSD	AEKAADKQPV
101	ADKADEVEEK	AGEPEREEPD	GQAVRKKALT	EEREQTVREK	AQKKDAETVK
151	KQAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSGS	IEKARSAAAK
201	EVQKMKTS DK	AEATHYLQMG	AYADRQSAEG	QRAKLAILGI	SSKVVGYQAG
251	HKTLYRVQSG	NMSADAVKKM	QDELKKHEVA	SLIRSIESK*	

Computer analysis of this amino acid sequence gave the following results:

- 15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF65 (SEQ ID NO: 382) shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) (SEQ ID NO: 386) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF65a nucleotide sequence [<SEQ ID 385>] (SEQ ID NO: 385) is:

40

1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCTGT	CCGGTTTTTT
51	CTTCGGTTTG	ATACTGGCGA	CGGTCAATTAT	TGCCGTATT	TTGTTTATATC
101	TGAACCAAG	CGGTCAAAAT	GCGTTCAAA	TCCCGTTCC	GTCTGAAGCAT
151	CCTGCAGAAA	CGGAAATCCT	GAAACCGAAA	AACCAAGCTA	AGGAATGACAT
201	CCAACCTGAA	CCGGCCGATC	AAAACGCTT	GTCCGAACCG	GATGCTGCGA

5
251 AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCCTT
301 GCCGACAAAG CCGACGAGGT TGAGGAAAAG GCGGACGAGC CGGAGCGGGA
351 AAAGTCGGAC GGACAGGCAG TGCGCAAGAA AGCACTGACG GAAGAGCGTG
401 AACAAACCGT CGGGGAAAAA GCGCAGAAGA AAGATGCCGA AACGGTTAAA
451 AAACAAGCGG TAAAACCATC TAAAGAAACA GAGAAAAAAG CTTCAAAAGA
501 AGAGAAAAAG GCGGAGAAGG AAAAAGTTGC ACCCAAACCG ACCCCGGAAC
551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCTGCCCAA
601 GAAGTGCAGA AAATGAAAAC GCCCGACAAG GCGGAAGCAA CGCATTATCT
10 651 GCAATGCGC GCGTATGCCG ACCGCCGAG CGCGGAAGGG CAGCGTGCCA
701 AACTGGCAAT CTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGGCGGGA
751 CATAAAACGC TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT
801 GAAAAAATG CAGGACGAGT TGAAAAACA TGAAGTCGCC AGCCTGATCC
851 GTTCTATCGA AAGCAAATAA

15 This encodes a protein having amino acid sequence [<SEQ ID 386>] (SEQ ID NO: 386):

20
1 MFMNKFSQSG KGLSGFFFGL ILATVIIAGI LFYLNQSGQN AFKIPVPSKQ
51 PAETEILPKP NQPKEDIQPE PADQNALSEP DAAKEAEQSD AEKAADKQPV
101 ADKADEVEEK ADEPEREKSD QQAVRKKALT EEREQTVGEK AQKKDAETVK
151 KQAVKPSKET EKKASKEEKK AEKEKVAPKP TPEQILNSGS IEKARSAAAK
201 EVQKMKTDPK AEATHYLQMG AYADRRSAEG QRAKLAILGI SSKVVGYYQAG
251 HKTLYRVQSG NMSADAVKKM QDELKKHEVA SLIRSIESK*

ORF65a (SEQ ID NO: 386) and ORF65-1 (SEQ ID NO: 384) show 96.5% identity in 289 aa overlap:

25
orf65a.pep 10 20 30 40 50 60
MFMNKFSQSGKGLSGFFFGLILATVIIAGILFYLNQSGQNAFKIPVPSKQPAETEILPKP
orf65-1 MFMNKFSQSGKGLSGFFFGLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILPKP
10 20 30 40 50 60
30
orf65a.pep 70 80 90 100 110 120
NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADEPEREKSD
|||:|
orf65-1 NQPKEDIQPEPADQNALSEPDAAEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP
70 80 90 100 110 120
35
orf65a.pep 130 140 150 160 170 180
GQAVRKKALTEEREQTVGEKAQKKDAETVKKQAVKPSKETEKASKEEKKAEKEKVAPKP
|||:|
orf65-1 GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKASKEEKKAAKEKVAPKP
130 140 150 160 170 180
40
orf65a.pep 190 200 210 220 230 240
TPEQILNSGSIEKARSAAAKEVQKMKTDPKAEATHYLQMGAYADRRSAEGQRAKLAILGI
|||:|
orf65-1 TPEQILNSGSIEKARSAAAKEVQKMKTSDKAEATHYLQMGAYADRQSAEGQRAKLAILGI
190 200 210 220 230 240
45
orf65a.pep 250 260 270 280 290
SSKVVGYYQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX
|||:|
orf65-1 SSKVVGYYQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX
250 260 270 280 290

Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 (SEQ ID NO: 382) shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) (SEQ ID NO: 388) from *N. gonorrhoeae*:

```

5      30      40      50      60      70      80
ORF65ng IIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLKNQPKEDIQPEPADQNALSEPDVAKE
ORF65      ILKPHNQLKEDIQPD PADQNALSEPDAATE
              10      20      30

10     90     100     110     120     130     140
ORF65ng AEQSDAEKAADKQPVADKADEVEEEKAGEPEREEPDGQAVRKKALTEEREQTVREKAQKKD
ORF65      AEQSDAENAADKQPVADKADEVEEEKAGEPEREEPDGQAVRKKALTEEREQTVREKAQKKD
              40      50      60      70      80      90

15     150     160     170     180     190     200
ORF65ng AETVKKKAVKPSKETTEKKASKEEKAAKEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM
ORF65      AETVKIQAVKPSKETTEKKASKEEKAAKEKVAPKPTPEQILNSGSIEKARSAAAKEVQKM
              100     110     120     130     140     150

20     210     220     230     240     250     260
ORF65ng KNFGQGSQRRIICKWARMPNPGARKGSPVNWQSWAYLPKWSAIRRDIKRFTACKAAICPP
ORF65      XNVRQGSXRRIICKWARMPNPGARKGSPVNWQSWAYLPKWSAIRRDIKRFTGCKAAICLP
              160     170     180     190     200     210

25     ORF65ng      MR
ORF65      ||
ORF65      MR

```

An ORF65ng nucleotide sequence [<SEQ ID 387>] (SEQ ID NO: 387) was predicted to encode a protein having amino acid sequence [<SEQ ID 388>] (SEQ ID NO: 388):

```

30      1  MFMNKFSQSG KGLSGFFFL ILATVVIAGI LLYLNQGGQN AFKIPAPSKQ
51  PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV
101 ADKADEVEEK AGEPEREEPD GQAVRKKALT EEREQTVREK AQKKDAETVK
151 KKA VKPSKET EKKASKEEK AAKEKVAPKP TPEQILNSRS IEKARSAAAK
201 EVQKMKNFGQ GGSQRRIICKW ARMPNPGARK GSPVNWQSWA YLPKWSAIRR
35  251 DIKRFTACKA AICPPMR*

```

After further analysis, the complete gonococcal DNA sequence [<SEQ ID 389>] (SEQ ID NO: 389) was found to be:

```

40      1  ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTCTT
51  CTTCGGTTTG ATACTGGCAA CGGTCATTAT TGCCGGTATT TTGCTTTATC
101 TGAACCAGGG CGGTCAAAT GCGTTCAAAA TCCCGGCTCC GTCGAAGCAG
151 CCTGCAGAAA CGGAAATCCT GAAACTGAAA AACCAGCCTA AGGAAGACAT
201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGTTGCGA
251 AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT

```

5	301	GCCGACAAag	cgcacgAGGT	TGAAGAAAag	GcGGgcgAgc	cggaACGGga
	351	aGAGCCGGAC	ggACAGGCAG	TGCGCAAGAA	AGCACTGAcg	gAAGAgcGTG
	401	AACAAACcgt	cagggAAAAA	GCGCagaaga	AAGATGCCGA	AACGgTTAAA
	451	AAacaaGCg	tAaaaccgtc	tAAAGAAACa	gagaaaaaag	cTtcaaaaga
	501	agagaaaaag	gcggcgaaag	aaaAAGttgc	acccaaacgc	accccggaac
10	551	aaatcctcaa	cagccgCagc	atcgaaaaag	cgcgtagtgc	cctgtgcaaa
	601	gaAgtgcaGA	AAatgaaaaa	ctTtgggcaa	ggcgGaagcc	aacgcattaT
	651	CTGcaaatgg	gcgcgtatgc	cgaccgtccg	gagcgcggaA	gggcagcgtg
	701	ccaaACtggc	aAtcttgGgc	atatctTccg	aagtgggtcgG	CTATCAGGCG
	751	GGACATAAAA	CGCTTTACCG	CGTGCAaagc	GGCAatatgt	ccgccgatgc
	801	gGTGAAAAAA	ATGCAGGACG	AGTTGAAAAA	GCATGGGGtt	gcCAGCCTGA
	851	TCCGTGcqAT	TGAAGGCAAA	TAA		

This encodes the following amino acid sequence [<SEQ ID 390>] (SEQ ID NO: 390):

15	1	MFMNKFSSQG	KGLSGFFFL	ILATVIIAGI	LLYLNQGGQN	AFKIPAPSKQ
	51	PAETEILKLK	NQPKEDIQPE	PADQNALSEP	DVAKEAEQSD	AEKAADKQPV
	101	ADKADEVEEK	AGEPEREEDP	GQAVRKKALT	EEREQTVREK	AQKKDAETVK
	151	KQAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
20	201	EVQKMKNFQG	GGSQRIICKW	ARMP TVRSAP	GQRAKLAILG	ISSEVVGYQA
	251	GHKTLYRVQS	GNMSADAVKK	MQDELKKHGV	ASLIRAIEGK	*

ORF65ng-1 (SEQ ID NO: 390) and ORF65-1 (SEQ ID NO: 384) show 89.0% identity in 290 aa overlap:

			10	20	30	40	50	60
25	orf65-1.pep	MFMNKFSQSGKGLSGFFFFGLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK						
	orf65ng-1	MFMNKFSQSGKGLSGFFFFGLILATVIIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLK						
			10	20	30	40	50	60
			70	80	90	100	110	120
30	orf65-1.pep	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP						
	orf65ng-1	NQPKEDIQPEPADQNALSEPDVAKEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP						
			70	80	90	100	110	120
			130	140	150	160	170	180
35	orf65-1.pep	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEEKKASKEEKKAAKEKVAPKP						
	orf65ng-1	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEEKKASKEEKKAAKEKVAPKP						
			130	140	150	160	170	180
			190	200	210	220	230	239
40	orf65-1.pep	TPEQILNSGSIEKARSAAAEVQKMKTSDKAEATHYL-QMGAYADRQSAEGQRAKLAILG						
	orf65ng-1	TPEQILNSRSIEKARSAAAEVQKMNFGQGSQRILCKWARMPTVRSAGQRAKLAILG						
			190	200	210	220	230	240
			240	250	260	270	280	290
45	orf65-1.pep	ISSKVVGYYQAGHKTLRYVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX						
	orf65ng-1	ISSEVVGYQAGHKTLRYVQSGNMSADAVKKMQDELKKHGVASLIRIEGKX						
			250	260	270	280	290	

On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 46

- 5 The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 391>] (SEQ ID NO: 391):

```

1  ATGAACCACG ACATCACTTT CCTCACCCCTG TTCCTACTCG GTkTCTTCGG
51  CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GcGTTTGs.s
101 TCCAAC TCCC CCCGCATATC AACCGCTTTT GGCTGATCCT GCTGCTTAAC
151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCCTGATAC TCGGATTAAT
201 CGGACAGGTC GCGGTTTCAC TCGACCAaAC CCGCGTCCTG CAGAATATTT
251 TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
301 GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAaATCGGCA AACCGATATG
351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
15 401 CCGCCTGCCT tGCGgTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AgCGGTAGTG CGGCAACGGG
501 CGGGTTATAT ATGCTTGCC T TGCACTGGG TACGCTGCCC AATCTttTAG
551 CAATCGGCAT TTTtTCCCTG CAACTGAaAwA AAATCATGCA AAACCGATAT
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
20 651 TGCCGTCCTG TGGCTGTAA

```

This corresponds to the amino acid sequence [<SEQ ID 392; ORF103>] (SEQ ID NO: 392; ORF103):

```

25 1  MNHDITFLTL FLLGXFGGTH CIGMCGGLSS AFXXQLPPhi NrfWLILLN
51  TGRVSSYTAI GLILGLIGQV GVS LDQTRVL QNilyTAANL LLLFLGLYLS
101 GISSLAakIE KIGKPIWRNL NPILNRLlPI KSIPAClAVG ILWGWLPcGL
151 VYSASLYALG SGSAATGGLY MLAFALGTLP NLLATGIFSL QLXKIMQnRY
201 IRLCTGLSVS LWALWKLAVL WL*

```

- 30 Further work elaborated the DNA sequence [<SEQ ID 393>] (SEQ ID NO: 393) as:

```

35 1  ATGAACCACG ACATCACTTT CCTCACCCCTG TTCCTACTCG GTTTCTTCGG
51  CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGCGC
101 TCCAAC TCCC CCCGCATATC AACCGCTTTT GGCTGATCCT GCTGCTTAAC
151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCCTGATAC TCGGATTAAT
201 CGGACAGGTC GCGGTTTCAC TCGACCAAAC CCGCGTCCTG CAGAATATTT
251 TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
301 GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAAATCGGCA AACCGATATG
351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
40 401 CCGCCTGCCT TGCGGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG
501 CGGGTTATAT ATGCTTGCC T TGCACTGGG TACGCTGCCC AATCTTTTAG
551 CAATCGGCAT TTTTTCCTG CAACTGAAAA AAATCATGCA AAACCGATAT
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
45 651 TGCCGTCCTG TGGCTGTAA

```

This corresponds to the amino acid sequence [[SEQ ID 394](#); [ORF103-1](#)] ([SEQ ID NO: 394](#); [ORF103-1](#)):

5

1	MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPhi	NRFWLILLLN
51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPi	KSIPACLAvg	ILWGWLPcGL
151	VYSASLYALG	SGSAATGGLY	MLAFALGTLP	NLLAIGIFSL	QLKKIMQnRY
201	IRLCTGLSVS	LWALWKLAVL	WL*		

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 (SEQ ID NO: 392) shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) (SEQ ID NO: 396) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
15	orf103.pep	MNHDITFLTFLFLGXFGGTHCIGMCGGLSSAFXQLPPHINRFWLILLNTGRVSSYTAI					
	orf103a	MNXDITFLTFLFLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLILLNTGRVSSYTAI					
		10	20	30	40	50	60
20	orf103.pep	GLILGLIGQVGVS LDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
	orf103a	GLILGLIGQVGVS LDQTRVXQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
25	orf103.pep	NPILNRLLPKISIPACLA VGILGWGWP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
	orf103a	NPILNRLLPKISIPACLA VGILGWGWP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
30	orf103.pep	NLLAIGIFSLQLXKIMQNR YIRLCTGLSVSLWALWKLAVLWLX					
	orf103a	NLXAIGIFSLQLXKIMQNR YIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

The complete length ORF103a nucleotide sequence [[<SEQ ID 395>](#)] ([SEQ ID NO: 395](#)) is:

35	1	ATGAACCANG	ACATCACTTT	CCTCACCCGT	TTCCTACTCG	GTTTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
	101	TCCAATCCC	CCCGCATATC	AACCGCTTNT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGATTAAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAAAC	CCGCGTCNTG	CAGAATATTT
40	251	TATACACGGC	CGCCAACTTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTTCCT	CCTTGCGGGC	AAAAATCGAG	AAAAATCGGA	AACCGATATG
	351	CGCGGAACCTG	AACCCGATAC	TCAACCGGCT	GTTATCCCAT	AAATCCATAC
	401	CCGCCGTGCTT	TGCGGTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTA
	451	GTTTACACGG	CGTCGCTTTA	CGCGCTGGGA	AGCGGTAGTG	CGGCAACGGG

501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTTNGG
551 CAATCGGCAT TTTTCCCTG CAACTGNAAA AAATCATGCA AAACCGATAT
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
651 TGCCGTCCTG TGGCTGTAA

This encodes a protein having amino acid sequence [<SEQ ID 396>] (SEQ ID NO: 396):

1	MNXDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPHI	NRXLWILLLN
51	TGRVSYTAI	GLILGLIGQV	GVSLDQTRVX	QNILYTAANL	LLFLGLGLYS
101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPI	KSIPACLAVG	ILGWGLPCGL
151	VYSASLYALG	SGSATGGGLY	MLAFALGTLF	NLXAIGIFSL	QLXKIMQNRY
201	IRLTGLSVS	LWLAKLAVL	WF*		

ORF103a (SEQ ID NO: 396) and ORF103-1 (SEQ ID NO: 394) show 97.7% identity in 222 aa overlap:

	10	20	30	40	50	60
orf103a.pep	MNXDITFLT	LFLFGFFGG	THCIGMCGGL	SSAFALQLP	PHINRXWL	LILLNTGRVSSYTAI
orf103-1	MNHDITFLT	LFLFGFFGG	THCIGMCGGL	SSAFALQLP	PHINRFWL	LILLNTGRVSSYTAI
	10	20	30	40	50	60
	70	80	90	100	110	120
orf103a.pep	GLILGLIGQ	VGVSLDQTR	VXQNILYTA	ANLLLLFL	GLYLSGI	SSLAAKIEKIGKPIWRNL
orf103-1	GLILGLIGQ	VGVSLDQTR	VLQNILYTA	ANLLLLFL	GLYLSGI	SSLAAKIEKIGKPIWRNL
	70	80	90	100	110	120
	130	140	150	160	170	180
orf103a.pep	NPILNRLLP	IKSIPACLA	VGILGWG	WLP	CGLVYSAS	LYALGSGSAATGGGLYMLAFALGTLP
orf103-1	NPILNRLLP	IKSIPACLA	VGILGWG	WLP	CGLVYSAS	LYALGSGSAATGGGLYMLAFALGTLP
	130	140	150	160	170	180
	190	200	210	220		
orf103a.pep	NLXAIGIF	SLQLXKIM	QNRYIRL	CTGLSV	SLWALW	KLAVLWLX
orf103-1	NLLAIGIF	SLQLKKIM	QNRYIRL	CTGLSV	SLWALW	KLAVLWLX
	190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 (SEQ ID NO: 392) shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) (SEQ ID NO: 398) from *N. gonorrhoeae*:

orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXQLPPHINRFWLILLLNTGRVSSYTAI	60
orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLLNTGRISSYTAI	60
orf103.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
	: : : :	
orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTASNLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120

5	orf103.pep	NPILNRLLPKISIPACLAVGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP	180
	orf103ng	NPILNRLLPKISIPACLAVGILWGWLPCGLVYSASLYALGSGSATTGGLYMLAFALGTLP	180
	orf103.pep	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWL	222
	orf103ng	NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence [[<SEQ ID 397>](#)] ([SEQ ID NO: 397](#)) is:

10	1	ATGAACCACG	ACATCACTTT	CCTCACCTG	TTCCTGCTCG	TTTTCTTCGG
	51	CGGAATCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
15	101	TCCAATCCC	CCCGCATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
	151	ACAGGACGGA	TAAGCAGCTA	TACGGCAATC	GGCCTGATGC	TCGGATTAAT
20	201	CGGACAATC	GGCATTTTAC	TCGACCAAAc	ccgcgTCCTG	CAAAATATTT
	251	tatacacagc	ctccaaCCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
25	301	GGTATTTCTT	CCTTGCGCGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
	351	GCGCAACCTG	AACCCGATAC	TCAACCGGCT	GCTGCCCATA	AAATCCATAC
30	401	CCGCCTGCCT	TGCTGTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGCG	CATCACTTTA	CGCGCTGGGA	AGCGGTAGTG	CGACAACCGG
35	501	CGGACTGTAT	ATGCTTGCCT	TTGCACTGGG	TACGCTGCCC	AATCTTTTGG
	551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
40	601	ATCCGCCTGT	GTACAGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
	651	TGCCGTCCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence [[<SEQ ID 398>](#)] ([SEQ ID NO: 398](#)):

25	1	MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPhi	NRFWLILLN
	51	TGRISSTAI	GLMLGLIGQL	GISLDQTRVL	QNILYTASNL	LLLFLGLYLS
30	101	<u>GISSLA</u> AKIE	KIGKPIWRNL	NPILNRLLPi	<u>KSIPACLA</u> VG	<u>ILWGWLPCGL</u>
	151	<u>VYSASLYALG</u>	<u>SGSATTGGLY</u>	<u>MLAFALGTL</u> P	<u>NLLAIGIFSL</u>	<u>QLKKIMQNRY</u>
35	201	IRLCTGLSVS	LWALWKLAVL	WL*		

In addition, ORF103ng ([SEQ ID NO: 398](#)) and ORF103-1 ([SEQ ID NO: 394](#)) show 97.3% identity in 222 aa overlap:

35	orf103-1.pep	10	20	30	40	50	60
	orf103ng	10	20	30	40	50	60
40	orf103-1.pep	70	80	90	100	110	120
	orf103ng	70	80	90	100	110	120
45	orf103-1.pep	130	140	150	160	170	180
	orf103ng	130	140	150	160	170	180
50	orf103-1.pep	190	200	210	220		
	orf103ng	190	200	210	220		

orf103ng
 |||||
 NLLAIGIFSLQLKKIMQNRRIYRLCTGLSVSLWALWKLAVLWLX
 190 200 210 220

- 5 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

- 10 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 399>] (SEQ ID NO: 399):

1 ATGGA^{AA}AACC AAAGGCCGCT CCTAGGCTTT CGCTTGGCAC TTTTGGCGGC
 51 GATGACGTGG GGAACGCTGC CGAT.TCCGT GCGGCAGGTA TTGAAGTTTG
 101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GGCGGCGGTA
 15 151 TTGTTTGT^{TT} TGCTGGCACT GGGCGGGCGG CTGCCGAAGC GGCGGGGATT
 201 TTTCTTGGTG CTCATTCAGG CTGCTGCTGC TCGGCGTGGC GGGCATTTCG
 251 GCAAAC^{TT}TG TGCTGATTGC CCAAGGGCTG CATTATATTT CGCCGACCAC
 301 GACGCAGGTT TTGTGGCAGA TTTCCGCCGT TACGATGATT GTwGTCTGGTG
 351 TGT^{TT}GGTGTT TAAAGACCGG ATGACTGCCG CTCAGAA^{AA}T CGGCTTGGTT
 20 401 TTGCTGCTTG CCGGT^{TT}GCT TATGTATTTT AACGATAAAT TCGGCGAGTT
 451 GTCGGGTTTG GGCGCGTATG C.AAGGGCGT GTTGTCTGTG GCGGCAGGCA
 501 GTATGGCATG GGTGTGTAAT GCCGTGGCGC AAAAGCTGCT GTCGGCGCAA
 551 TTCGGGCGCG AACAGATTCT GCTGTTGATT TATGCGGCAA GTGCCGCCGT
 601 GTTCCTGCCG TTTGCCGAAC CGGCACACAT CGGAAGTATG GACGGTACGT
 25 651 TGGCGTGGGT ATGTATTGCG TATTGCTGCT TGAATACGTT AATCGGTTAC
 701 GGCTCGTT^{CG} GCGAGGCGTT GAAACATTGG GAGGCTTCCA AAGTCAGCGC
 751 GGTAACAACC TTGCTCCCCG TGTTTACCGT AATAAATACT TTGCTCGGGC
 801 ATTATGTGAT GCCTGAAACT TTTGCCGCGC CGGA..

- 30 This corresponds to the amino acid sequence [<SEQ ID 400; ORF104>] (SEQ ID NO: 400; ORF104):

1 MENQRPLLGF RLALLAAMTW GTLPXSVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFLVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MYFNDKFGEL
 35 151 SGLGAYXKGV LLCAAGSMAW VCN^{AV}AQKLL SAQFGPQQIL LLIYAASAAV
 201 FLPPFAEPAHI GSMDGTLAWV CIAYCCLNTL IGYGSFGEAL KHWEASKVSA
 251 VTLLLPVFTV INTLLGHYVM PETFAAP...

Further work revealed further partial DNA sequence [<SEQ ID 401>] (SEQ ID NO: 401):

40 1 ATGGA^{AA}AACC AAAGGCCGCT CCTAGGCTTC GCGTTGGCAC TTTTGGCGGC
 51 GATGACGTGG GGAACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
 101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GGCGGCGGTA
 151 TTGTTTGT^{TT} TGCTGGCACT GGGCGGGCGG CTGCCGAAGC GGCGGGATT
 201 TTCTTGGTGC TCATTCAGGC TGCTGCTGCT CGGCGTGGCG GGCATTTCGG
 45 251 CAAAC^{TT}TGT GCTGATTGCC CAAGGGCTGC ATTATATTTT GCCGACCACG

5 301 ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATTG TTGTCGGTGT
 351 GTTGGTGT TT AAAGACCGGA TGA CTGCCGC TCAGAAAATC GGCTTGGTTT
 401 TGCTGCTTGC CGGTTTGCTT ATGTTTTTTA ACGATAAATT CGGCGAGTTG
 451 TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CCGCAGGCAG
 501 TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCGCAAT
 551 TCGGGCCGCA ACAGATTCTG CTGTTGATT ATGCGGCAAG TGCCGCCGCTG
 601 TTCCTGCCGT TTGCCGAACC GGCACACATC GGAAGTTTGG ACGGTACGTT
 651 GCGGTGGGTT TGTTTTGCGT ATTGCTGCTT GAATACGTTA ATCGGTTACG
 10 701 GCTCGTTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATAwTwCTT TGCTCGGGCA
 801 TTATGTGATG CCTGAAACTT TTGCCGCGCC GGA...

This corresponds to the amino acid sequence [<SEQ ID 402; ORF104-1>] (SEQ ID NO: 402; ORF104-1):

15 1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 L FVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
 201 FLPPFAEPAHI GSLDGT LAWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
 20 VTTLLPVFTV IXXLLGHYVM PETFAAP...

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein (SEQ ID NO: 1138) of *H. influenzae* (accession number U32769)

25 ORF104 (SEQ ID NO: 400) and HI0878 (SEQ ID NO: 1138) show 40% aa identity in 277aa overlap:

30 orf104 4 QRPLLGFRLALLAAMTWGTLPSVRQVLKFVDAPTLVXXXXXXXXXXXXXXXXXXXXP- 62
 Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P
 HI0878 3 QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSL LALLAYKKQLPE 62

30 orf104 63 --KRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 120
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
 HI0878 63 LMKVRQYAW---IMLIGVIGLTSNFLFSSSLNIEPSVAQIFIHLSSFGMLICGVLIF 118

35 orf104 121 KDRMTAAQKIXXXXXXXXXXXMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180
 K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+
 HI0878 119 KEKLGLHQKIGLFLLLIGLGLFFNDRFADFAGLNQYSTGVILGVGGALI WVAYGMAQKLM 178

orf104 181 SAQFGPQQILLIYAASAAVFLPPFAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL 240
 +F QQILL++Y A F+P A+ + + + LA +C YCCLNTLIGYGS+ EAL
 HI0878 179 LRKFNSQQILLMMYLGCAIAFMPMADFSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237

40 orf104 241 KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277
 W+ SKVS V TL+P+FT++ + + HY P FAAP
 HI0878 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAP 274

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF104 (SEQ ID NO: 400) shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) (SEQ ID NO: 404) from strain A of *N. meningitidis*:

5	orf104.pep	10 20 30 40 50 60	MENQRPLLGFRLALLAAMTWGTLPSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
	orf104a	10 20 30 40 50 60	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
10	orf104.pep	70 80 90 100 110 120	LPKRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIWVGVLVF
	orf104a	70 80 90 100 110 120	LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIWVGVLVF
15	orf104.pep	130 140 150 160 170 180	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL
	orf104a	130 140 150 160 170 180	KDRMTAAQKIGLVLLLAGLLMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
20	orf104.pep	190 200 210 220 230 240	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYSFGEAL
	orf104a	190 200 210 220 230 240	SAQFGPQQIILLIYAASAAVFLPFAELAHIGSLDGTALAWVCFAYCCLNTLIGYSFGEAL
25	orf104.pep	250 260 270	KHWEASKVSAVTLLPVFTVINTLLGHYVMPETFAAP
	orf104a	250 260 270 280 290 300	KHWEASKVSAVTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGYAGALVVVGGAVTAAVG

The complete length ORF104a nucleotide sequence [<SEQ ID 403>] (SEQ ID NO: 403) is:

30	1	ATGGA	AAACC	AAAGG	CCGCT	CCTAG	GCTTC	GCGTT	TGGCAC	TTTGG	CGGC
	51	GATGA	CGTGG	GGAAC	GCTGC	CGATT	GCCGT	GCGGC	AGGTA	TTGAA	AGTTT
	101	TCGAT	GCGCC	GACGC	TGGTG	TGGGT	GCGTT	TTACCG	TGGC	GGCGG	CGGTA
	151	TTGTT	TGTTT	TGCTG	GCATT	GGGCG	GGCGG	CTGCC	GAAAT	GGCGG	ATTT
35	201	TTCTT	GGTGC	TCATT	CAGGC	TGCTG	GCTGCT	CGGCG	TGGCG	GGCAT	TTTCG
	251	CAAAC	TTTGT	GCTGA	TGCC	CAAGG	GCTGC	ATTAT	ATTTT	GCCGA	CCACG
	301	ACGCA	GGTTT	TGTGG	CAGAT	TTCCG	CGTTT	ACGAT	GATTG	TTGTC	GGTGT
	351	GTTGT	GTGTT	AAAGA	CCGGA	TGACT	GCCGC	TCAGA	AAATC	GGCTT	GGTTT
	401	TGCTG	CTTGC	CGGTT	TGCTT	ATGTT	TTTTTA	ACGATA	AAAT	CGGCG	AGTTG
	451	TCGGG	TTTGG	GCGCG	TATGC	GAAGG	GCGTG	TTGCT	GTGTG	CGGCAG	GACAG
40	501	TATGG	CATGG	GTGTT	TATG	CCGTG	GCGCA	AAAGC	TGCTG	TCGCG	CAAT
	551	TCGGG	CCGCA	ACAGA	TCTG	CTGTT	GATTT	ATGCG	GCAAG	TGCCG	CGTG
	601	TTCTG	CCCGT	TTGCC	GAACT	GGCAC	CACATC	GGAAG	TTTGG	ACGGT	ACGTT
	651	GGCGT	TGGGT	TGTTT	TGCGT	ATTGT	GCTGT	GAATAC	GTTA	ATCGT	TACG
	701	GCTCG	TTCGG	CGAGG	CGTTG	AAACA	TGGG	AGGCT	TCCAA	AGTCA	GCGCG
45	751	GTAACA	AACT	TGCTC	CCCGT	GTTTAC	CGTA	ATATTT	TCTT	TGCTC	GGGCA
	801	TTATG	TGATG	CCTGA	TA	TTGCC	GCGCC	GGATAT	GAAAC	GGTTT	TGGGT
	851	ATGCC	GCGCG	ACTGT	GCTG	GTCGG	GGGTG	CGGTT	ACGGC	GGCGG	TGGGG
	901	GACAG	GCTGT	TCAAAC	GCCG	CTAG					

50 This encodes a protein having amino acid sequence [<SEQ ID 404>] (SEQ ID NO: 404):

1 MENQRPLLGF ALALLAAMTW GTLPPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFVLLALGGR LPKWRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
 201 FLPFAELAHIGSLDGTALWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
 251 VTLLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYAGALVV VGGAVTAAVG
 301 DRLFKRR*

ORF104a (SEQ ID NO: 404) and ORF104-1 (SEQ ID NO: 402) show 98.2% identity in 277 aa overlap:

		10	20	30	40	50	60
orf104a.pep		MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
orf104-1		MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf104a.pep		LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF					
orf104-1		LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf104a.pep		KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
orf104-1		KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf104a.pep		SAQFGPQQIILLIYAASAAVFLPFAELAHIGSLDGTALWVCFAYCCLNTLIGYGSFGEAL					
orf104-1		SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSLDGTALWVCFAYCCLNTLIGYGSFGEAL					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf104a.pep		KHWEASKVSAVTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYAGALVVGGAVTAAVG					
orf104-1		KHWEASKVSAVTLLPVFTVIXLLGHYVMPETFAAP					
		250	260	270			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF104 (SEQ ID NO: 400) shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) (SEQ ID NO: 406) from *N. gonorrhoeae*:

orf104.pep	MENQRPLLGFRLALLAAMTWGTLPSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
orf104ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF	120
orf104ng	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF	120

orf104.pep KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180
 orf104ng KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 180

5 orf104.pep SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL 240
 orf104ng SAQFGPQQIILLIYAASAAVFLXAEPAHIGSLDGTALAWCFVYCCLNTLIGYGSFGEAL 240

10 orf104.pep KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277
 orf104ng KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVVGAVTAAVG 300

The complete length ORF104ng nucleotide sequence [SEQ ID 405] (SEQ ID NO: 405) is predicted to encode a protein having amino acid sequence [SEQ ID 406] (SEQ ID NO: 406):

1 MENQRPLLGF ALALLAAMTW GTLPPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFVLLALGGR LPKRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLF KDRMTAAQKI GLVLLLVGLL MFFNDKFGEL
 151 SGLGAYAGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
 201 FLLXAEPAHI GSLDGTALAW CFVYCCLNTL IGYGSFGEAL KHWEASKVSA
 251 VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG
 301 DRPFKRR*

Further work revealed the complete gonococcal nucleotide sequence [SEQ ID 407] (SEQ ID NO: 407):

1 ATGGAACACC AAAGGCCGCT CCTAGGCTTC GCGTTGGCAC TTTTGGCGGC
 51 GATGACGTGG GGGACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
 101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GGCGGCGGTA
 151 TTGTTTGT TTGCTGGCATT GGGCGGGCGG CTGCCGAAGC GGCGGGATT
 201 TTCTTGGCAT TCATTCAGGC TGCTGCTGCT CGGCGTGACG GGCATTTCCG
 251 CAACTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTTT GCCGACCACG
 301 ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATTG TTGTCGGCGT
 351 GTTGGTGTTT AAAGACCGGA tgaCTGCCGC GCAGAAAATC GGTTTGGTTT
 401 TGCTGcttgT CCGTttgCTT ATGTTTtTa ACGACAAATT CGGCGAGTTG
 451 TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CGGCAGGCAG
 501 TATGGCCTGG GTGTGTATG CCGTGCGCA AAAGCTGCTG TCGGCGCAAT
 551 TCGGGCCGCA ACAGATTCTG CTGTTGATTT ATGCGGcaag tgccgcccGTG
 601 TTCctgccgT TTGccgaaCC GGCACACATC GGAAGTTTgg aCGGTACgtt
 651 GGCGTGCGTT TGTTTTGTGT ATTGCTGCTT GAATACGTTA ATCGGTTACG
 701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATATTTTCTT TGCTCGGGCA
 801 TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTTTGGGTT
 851 ATGTCGCGCG ACTGGTCGTG GTCGGGGGTG CCGTTACGGC GGCGGTGGGG
 901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence [SEQ ID 408; ORF104ng-1] (SEQ ID NO: 408; ORF104ng-1):

1 MENQRPLLGF ALALLAAMTW GTLPPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFVLLALGGR LPKRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLF KDRMTAAQKI GLVLLLVGLL MFFNDKFGEL
 151 SGLGAYAGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
 201 FLPFAEPAHI GSLDGTALAW CFVYCCLNTL IGYGSFGEAL KHWEASKVSA

251 VTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG
301 DRPFKRR*

ORF104ng-1 (SEQ ID NO: 408) and ORF104-1 (SEQ ID NO: 402) show 97.5% identity in 277 aa overlap:

```

10      10      20      30      40      50      60
orf104-1.pep  MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAFLFVLLALGGR
orf104ng-1    MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAFLFVLLALGGR
              10      20      30      40      50      60

15      70      80      90      100     110     120
orf104-1.pep  LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTQVLWQISPTMIVVGVLVF
orf104ng-1    LPKRRDFSWHSFRLLLLGVGTGISANFVLIAQGLHYISPTTQVLWQISPTMIVVGVLVF
              70      80      90      100     110     120

20      130     140     150     160     170     180
orf104-1.pep  KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
orf104ng-1    KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
              130     140     150     160     170     180

25      190     200     210     220     230     240
orf104-1.pep  SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL
orf104ng-1    SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL
              190     200     210     220     230     240

30      250     260     270
orf104-1.pep  KHWEASKVSAVTLLPVFTVIXXLLGHYVMPETFAAP
orf104ng-1    KHWEASKVSAVTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVGGAVTAAVG
              250     260     270     280     290     300

```

In addition, ORF104ng-1 (SEQ ID NO: 408) shows significant homology with a hypothetical *H. influenzae* protein (SEQ ID NO: 1138):

```

35  gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306
    Score = 237 bits (598), Expect = 8e-62
    Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

Query: 30  QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXXP- 88
          Q+P          M WG+LPIA++QVL  ++A T+VW          P
Sbjct: 3   QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSL LALLAYKKQLPE 62

40  Query: 89  --KRRDFSWHSFRLLLLGVGTGISANFVLIAQGLHYISPTTQVLWQISPTMIVVGVLVF 146
          K R ++W    ++L+GV G+++NF+L +  L+YI P+  Q+   +S F M++ GVL+F
Sbjct: 63  LMKVRQYAW----IMLIGVIGLT SNFLLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLIF 118

Query: 147  KDRMTAAQKIXXXXXXXXXXXMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206
          K+++  QKI          +FFND+F  +GL Y+ GV+L  G++ WV Y +AQKL+
45  Sbjct: 119 KEKLGHLHQKIGLFLLLIGLGLFFNDRF DAFAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178

```

Query: 207 SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTALAWVCFVYCCNLTLIGYGSFGEAL 266
+F QQILL++Y A F+P A+ + + L LA +CF+YCCNLTLIGYGS+ EAL
Sbjct: 179 LRKFNSQQILLMMYLGCAIAFMPMADFSQVQELT-PLALICFIYCCNLTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMN 306
W+ SKVS V TL+P+FT++FS + HY P FAAP++N
Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 409](#)] ([SEQ ID NO: 409](#)):

```
15      1  ATGGTAGCTC  GTCGGGCTCA  TAACCCGAAG  GTCGTAGGTT  CGAATCCTGT
      51  .CCCGCAACC  TAATTTCAAA  CCCCTCGGTT  CAATGCCGAG  GG.GTTTTGT
101    T.TTGCTGT  TTCCTGTTTC  CTGTTTCCTG  CCGCCTCCGT  TTTTGCCGG
151    ATTTTCCTT  CGGCCGCAAT  ATCGGAACGG  CAGACCGCCG  TCTGTTTGCG
201    GTTGCAAATT  CAGGCAGTTT  GGCTACAATC  TTCCGCATTG  TCTTCAAGAA
20      251  AGCCAACCAT  GCCGACCGTC  CGTTTACCG  AATCCGTCAG  CAAACAAGAC
301    CTTGATGCTC  TGTTTCGAGTG  GGCAAAAGCA  AGTTACGGTG  CAGAAAAGTTG
351    CTGGAACACG  CTGTATCTGA  ACGGTCysCC  TTTGGGCAAC  CTGTCGCCGG
401    AATGGGTGGA  ACGCGTsmmA  AAAGACTGGG  AGGCAGGCTG  CyCGGAGTCT
451    TCAGACGGCA  TTTTTCGTGA  TgCGGACGGc  TgGcCtGATA  TGgGCGGAcg
25      501    cTTACAGCAC  CTCGCCCTCG  GTTGGCACTG  TGCGGGGCTG  TTGGACGgsT
551    GCGCAACGA  GTGTTTCGAC  CTGACCGACG  GCGGCGGCAA  CCCCTTGTTT
601    ACGCTCGaAc  GCGCCGyTTT  mCGTCCTkTC  GGACTGCTCA  GCCGCGCCGT
651    CCATCTCAAC  GGTCTGACCG  AATCGGACGG  CCGATGGCAT  TTCTGGATAG
701    GCAGGCGCAG  TCCGCACAAA  GCAGTCGATC  CCAACAAACT  CGACAATACT
30      751    rCCGCCGGCG  GTGTTTCGG  CGCGGAAATG  CCGTCTGAAG  CCGTGTGTCG
801    CGAAAGCAGC  GAAGAAGCCG  GTTTGGATAA  AACGCTGcTT  CCGCTCATCC
851    GCCCGGTATC  GCAGCTGCAC  AGCCTGCGCT  CCGTCAGCCG  GGGTGATAC
901    AATGAAATCC  TGTATGTATT  CGATGCCGTC  CTGCCG...
```

35 This corresponds to the amino acid sequence [[SEQ ID 410](#); [ORF105](#)] ([SEQ ID NO: 410](#); [ORF105](#)):

```
40      1  MVARRAHNP  K VVGSNPXPAT  XFQTPRFNAE  XVLXLPVSCF  LFPAASVFCR
      51  IFLPAAISER  QTAVCLRLQI  QAVWLQSSAL  SSRKPTMPTV  RFTESVSKQD
101    LDALFEWAKA  SYGAESCWKT  LYLNGXPLGN  LSPEWVERVX  KDWEAGCXES
151    SDGIFLNADG  WPDMMGRLQH  LALGWHCAGL  LDGWRNECFD  LTDGGGNPLF
201    TLERAXRXPX  GLLSRAVHLN  GLTESDGRWH  FWIGRRSPHK  AVDPNKLNDT
251    XAGGVSGGEM  PSEAVCRESS  EEAGLDKTL  PLIRPVSQLH  SLRSVSRGVH
301    NEILYVFDAV  LP...
```

45 Further work revealed the complete nucleotide sequence [[SEQ ID 411](#)] ([SEQ ID NO: 411](#)):

	1	ATGCCGACCG	TCCGTTTAC	CGAATCCGTC	AGCAACAAG	ACCTTGATGC
	51	TCTGTTTCGAG	TGGGCAAAAG	CAAGTTACGG	TGCAGAAAGT	TGCTGGAAAA
	101	CGCTGTATCT	GAACGGTCTG	CCTTTGGGCA	ACCTGTCGCC	GGAATGGGTG
5	151	GAACGCGTCA	AAAAAGACTG	GGAGGCAGGC	TGCTCGGAGT	CTTCAGACGG
	201	CATTTTCTCTG	AATGCGGACG	GCTGGCCTGA	TATGGGCGGA	CGCTTACAGC
	251	ACCTCGCCCT	CGGTTGGCAC	TGTGCGGGG	TGTTGGACGG	CTGGCGCAAC
	301	GAGTGTTTCG	ACCTGACCBA	CGGCGGCGGC	AACCCCTTGT	TCACGCTCGA
	351	ACGCGCCGCT	TTCCGTCCTT	TGCGACTGCT	CAGCCGCGCC	GTCCATCTCA
10	401	ACGGTCTGAC	CGAATCGGAC	GGCCGATGGC	ATTTCTGGAT	AGGCAGGCGC
	451	AGTCCGCACA	AAGCAGTCGA	TCCCAACAAA	CTCGACAATA	CTGCCGCCGG
	501	CGGTGTTTCC	GGCGGCGAAA	TGCCGTCTGA	AGCCGTGTGT	CGCGAAAGCA
	551	GCGAAGAAAG	CGGTTTGGAT	AAAACGCTGC	TTCCGCTCAT	CCGCCCGGTA
	601	TCGCAGCTGC	ACAGCCTGCG	CTCCGTCAGC	CGGGGTGTAC	ACAATGAAAT
15	651	CCTGTATGTA	TTGATGCCC	TCCTGCCCCA	AACCTTCTCTG	CCTGAAAAATC
	701	AGGATGGCGA	AGTGGCGGGT	TTTGAGAAAA	TGGACATCGG	CGGTCTGTTG
	751	GATGCCATGT	TGTCGGGAAA	CATGATGCAC	GACGCGCAAC	TGGTTACGCT
	801	GGACGCGTTT	TGCCGTTACG	GTCTGATTGA	TGCCGCCCAT	CCGCTGTCCG
	851	AGTGGCTGGA	CGGCATACGT	TTATAG		

20 This corresponds to the amino acid sequence [SEQ ID 412; ORF105-1] (SEQ ID NO: 412;
ORF105-1);

25

1	MPTVRFTESV	SKQDLLDALFE	WAKASYGAES	CWKTLYLNGL	PLGNLSP EW
51	ERVKKDWEAG	CSESSDGIFL	NADGW PDMGG	RLQHLALGWH	CAGLLDGWRN
101	ECFDLTDGGG	NPLFTLERA	FRPFGLLSRA	VHLNGLTESD	GRWHFWIGRR
151	SPHKAVDPNK	LDNTAAGGVS	GGEMPSEAVC	RESSEAEGLD	KTLLPLIRPV
201	SQHLSLRSVS	RGVHNEILYV	FDAVLPETFL	PENQDGEVAG	EKKMDIGLL
251	DAMLSGNMMH	DAQVLTDAF	CRYGLIDAAH	PLSEWLDGIR	L*

Computer analysis of this amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 (SEQ ID NO: 410) shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) (SEQ ID NO: 414) from strain A of *N. meningitidis*:

		60	70	80	90	100	110
35	orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALFEWAKASYGAES					
					:		
	orf105a	MPTVRFTESVSKHDLDALFEWAKASYGAES					
					10	20	30
		120	130	140	150	160	170
40	orf105.pep	CWKTLTYLNGXPLGNLSPEWVERVKDW EAGCXESSDGIFLNADGW PDMGRR LQH LALGWH					
		: :					
	orf105a	CWKTLTYLNLGPLGNLSPEWAERVKKDWEAGCSESSDGIFLNADGW PDMGRR LQH LARIWK					
		40	50	60	70	80	90
		180	190	200	210	220	230
45	orf105.pep	CAGLLDGWRNECFDLTDGGGNPLFTLERAXRXPGLLSRAVHLNGLTESDGRWHFWIGRR					
		: : :					
	orf105a	EAGLLHGWRDECFDLTDGGSNPLFALERAAFPPGILLSRAVHLNGLVESDGRWHFWIGRR					
		100	110	120	130	140	150

-330-

		240	250	260	270	280	290
orf105.pep		SPHKAVDPNKL	DNTXAGGV	SGGEMPSE	AVCRESSE	EAGLDTLL	PLIRPVSQLHSLRSVS
			:		:		
orf105a		SPHKAVDPDKL	DNTAAGGV	SSGELPSE	TVCRESSE	EAGLDTLL	PLIRPVSQLHSLRPVS
5		160	170	180	190	200	210
		300	310				
orf105.pep		RGVHNEILYV	FDAVL	P			
orf105a		RGVHNEILYV	FDAVL	PETFL	PENQDGE	VAGFEKMD	IGLLAAMLSGNMMHDAQLVTLDAF
10		220	230	240	250	260	270

The complete length ORF105a nucleotide sequence [[SEQ ID 413](#)] ([SEQ ID NO: 413](#)) is:

	1	ATGCCGACCG	TCCGTTTAC	CGAATCCGTC	AGCAAACACG	ACCTTGATGC
	51	CCTATTCGAG	TGGGCAAAGG	CAAGTTACGG	TGCGGAAAGT	TGCTGGAAAA
15	101	CGCTGTATCT	GAACGGTCTG	CCTTTGGGCA	ATCTGTCGCC	GGAATGGGCG
	151	GAGCGCGTCA	AAAAAGACTG	GGAGGCAGGC	TGCTCGGAGT	CTTCAGACGG
	201	CATTTTCCTG	AATGCGGACG	GCTGGCCAGA	TATGGGCAGA	CGCTTGCAGC
	251	ACCTCGCCCG	AATATGGAAA	GAAGCGGGAC	TGCTTCACGG	CTGGCGCGAC
	301	GAGTGTTCG	ACCTGACCGA	CGGCGGCAGC	AATCCCTTGT	TCGCGCTCGA
20	351	ACGCGCCGCT	TTCCGTCCGT	TCGGA	CTGCT	GTCCATCTCA
	401	ACGGTTTGGT	CGAATCGGAC	GGCCGATGGC	ATTTCTGGAT	AGGCAGGCGC
	451	AGTCCGCACA	AAGCAGTCGA	TCCCGACAAA	CTCGACAATA	CTGCCGCCGG
	501	CGGTGTTTCC	AGCGGTGAAT	TGCCGTCTGA	AACCGTGTGT	CGCGAAAGCA
	551	GCGAAGAAGC	CGGTTTGGAT	AAAACGCTGC	TTCCGCTCAT	CCGCCCGGTA
25	601	TCGCAGCTGC	ACAGCCTGCG	CCCCGTGAGC	CGGGGTGTGC	ACAATGAAAT
	651	CCTGTATGTA	TTCGATGCCG	TCCTGCCCGA	AACCTTCCTG	CCTGAAAATC
	701	AGGATGGCGA	AGTGGCGGGT	TTTGAGAAAA	TGGACATCGG	CGGTCTGTTG
	751	GCTGCCATGT	TGTCGGGAAA	CATGATGCAC	GACGCGCAAC	TGGTTACGCT
	801	GGACGCGTTT	TGCCGTTACG	GTCTGATTGA	TGCCGCCCAT	CCGCTGTCCG
30	851	AGTGGCTGGA	CGGCATACGT	TTATAG		

This encodes a protein having amino acid sequence [[SEQ ID 414](#)] ([SEQ ID NO: 414](#)):

	1	MPTVRFTESV	SKHDL	DALFE	WAKASYGAES	CWKTLYLNGL	PLGNL	SPEWA
	51	ERVKKDWEAG	CSESSD	GIFL	NADGW	PDMGR	RLQHL	LARIWK
35	101	ECFDLTDGGS	NPLFAL	ERAA	FRPFG	LLSRA	VHLNGL	VESD
	151	SPHKAVDPDK	LDNTA	AGGVS	SGELP	SETVC	RESSEE	EAGLD
	201	SQLHSLRPVS	RGVHNE	ILYV	FDAVL	PETFL	PENQD	GEVAG
	251	AAMLSGNMMH	DAQLV	TLDAF	CRYGL	IDAAH	PLSEW	LDGIR
								L*

40 ORF105a ([SEQ ID NO: 414](#)) and ORF105-1 ([SEQ ID NO: 412](#)) show 93.8% identity in 291 aa overlap:

		10	20	30	40	50	60
orf105a.pep		MPTVRFTESV	SKHDL	DALFE	WAKASYGAES	CWKTLYLNGL	PLGNL
45	orf105-1	MPTVRFTESV	SKQDL	DALFE	WAKASYGAES	CWKTLYLNGL	PLGNL
		10	20	30	40	50	60
		70	80	90	100	110	120
orf105a.pep		CSESSD	GIFLN	NADGW	PDMGR	RLQHL	ARIWKE
50	orf105-1	CSESSD	GIFLN	NADGW	PDMGR	RLQHL	ALGWHC
		70	80	90	100	110	120

-331-

		130	140	150	160	170	180
	orf105a.pep	FRPFGLLSRAVHLNGLVESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGELPSETVC					
	orf105-1	FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPNKLDNTAAGGVSGGEMPSEAVC					
5		130	140	150	160	170	180
	orf105a.pep	RESSEEAGLDKTLPLIRPVSQHLSLRPVSRGVHNEILYVFDAVLPEFLPENQDGEVAG					
	orf105-1	RESSEEAGLDKTLPLIRPVSQHLSLRSVSRGVHNEILYVFDAVLPEFLPENQDGEVAG					
10		190	200	210	220	230	240
	orf105a.pep	RESSEEAGLDKTLPLIRPVSQHLSLRPVSRGVHNEILYVFDAVLPEFLPENQDGEVAG					
	orf105-1	RESSEEAGLDKTLPLIRPVSQHLSLRSVSRGVHNEILYVFDAVLPEFLPENQDGEVAG					
		190	200	210	220	230	240
	orf105a.pep	FEKMDIGGLLAAMLSGNMMHDAQLVTLDACRYGLIDAAHPLSEWLDGIRLX					
	orf105-1	FEKMDIGGLLDAMLSGNMMHDAQLVTLDACRYGLIDAAHPLSEWLDGIRLX					
15		250	260	270	280	290	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 (SEQ ID NO: 410) shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) (SEQ ID NO: 416) from *N. gonorrhoeae*:

20	orf105.pep	MVARRAHNPKVVGSNPXPATXFQTPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER	60
	orf105ng	MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER	55
	orf105.pep	QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	120
	orf105ng	QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	115
25	orf105.pep	LYLNGXPLGNLSPEWVERVXKDWEAGCXESSDGI FLNADGWPDMMGRLQHLALGWHCAGL	180
	orf105ng	LYLNRLPLGNLSPEWAERIKKDWEAGCSESSNGI FLNADGWPDMMGRLQHLARTWNKAGL	175
	orf105.pep	LDGWRNECFDLTDGGGNPLFTLERAXRXPXGLLSRAVHLNGLTESDGRWHFWIGRRSPHK	240
30	orf105ng	LHGWRNECFDLTDGGGNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK	235
	orf105.pep	AVDPNKLDNTXAGGVSGGEMPSEAVCRESSEEAGLDKTLPLIRPVSQHLSLRSVSRGVH	300
	orf105ng	AVDPGKLDNIAGGVSGGEMPSEAVCRESSEEAGLDKTLFPLIRPVSRLHSLRPVSRGVH	295
35	orf105.pep	NEILYVFDAVL	312
	orf105ng	NEILYVFDAVLPEFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVTLDAFYRYG	355

A complete length ORF105ng nucleotide sequence [<SEQ ID 415>] (SEQ ID NO: 415) was predicted to encode a protein having amino acid sequence [<SEQ ID 416>] (SEQ ID NO: 416):

40	1	MVARRAHNPK	VVGSNPAPAT	KYQTPRFNAE	GVLFLLFPAA	SVFCRIFLPA
	51	AI SERQA AVC	LRLQIQAVWL	QSSALCSRKP	AMPTVRFTES	VSKQDLDALE
	101	ERAKASYGAE	SCWKTLYLNR	LPLGNLSPEW	AERIKKDEA	GCESSNGIF
	151	LNADGWPDMMG	GRLQHLARTW	NKAGLLHGWR	NECFDLTDGG	GNPLFTLERA

201 AFRPFGLLIR AVHLNGLVES NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV
 251 SGGEMPSEAV CRESSEEAGL DKTLPFLIRP VSRLHSLRPV SRGVHNEILY
 301 VFDAVLPETF LPENQDGEVA GF EKMDIGGL LDAMLSKNMM HDAQLVTLDA
 351 FYRYGLIDAA HPLSEWLDGI RL*

Further work revealed the complete nucleotide sequence [SEQ ID 417] (SEQ ID NO: 417):

1 ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC
 51 CCTGTTTCGAG CGGGCAAAAG CAAGTTACGG TGCCGAAAGT TGCTGGAAAA
 101 CGCTGTATCT GAACCGTCTT CCTTTGGGCA ATCTGTCGCC GGAATGGGCT
 151 GAGCGCATCA AAAAAGACTG GGAGGCAGGC TGCTCCGAGT CTTCAGACGG
 201 CATTTTTCTG AATGCGGACG GCTGGCCGGA TATGGGCGGA CGCTTGCAGC
 251 ACCTCGCCCG CACATGGAAC AAGGCGGGGC TGCTTCACGG ATGGCGCAAC
 301 GAGTGTTCG ACCTGACCGA CGGCGGCGGC AACCCCTTGT TCACGCTCGA
 351 ACGCGCCGCT TTCGTCCGT TCGGACTACT CAGCCGCGCC GTCCATCTCA
 401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC
 451 AGTCCGCACA AAGCAGTCGa tcCCGGCAAG CTCGACAATA TGCCGGCGCG
 501 CGGTGTTTCC GGCGGCGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCAGTA
 601 TCGCGGCTGC ACAGCCTTCG CCCCGTCAGC CGAGGTGTGC ACAATGAAAT
 651 CCTGTATGTG TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCTGAAAATC
 701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG
 751 GATGCCATGT TGTCGAAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
 801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence [SEQ ID 418; ORF105ng-1] (SEQ ID NO: 418; ORF105ng-1):

1 MPTVRFTESV SKQDLDALE RAKASYGAES CWKTLYLNL PLGNLSPEWA
 51 ERIKDWEG CSESSDGIFL NADGWPMGG RLQHLARTWN KAGLLHGWRN
 101 ECFDLTDGGG NPLFTLERA FRPFGLLSRA VHLNGLVESN GRWHFWIGRR
 151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEEAGLD KTLFPLIRPV
 201 SRLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
 251 DAMLSKNMMH DAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L*

ORG105ng-1 (SEQ ID NO: 418) and ORF105-1 (SEQ ID NO: 412) show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
orf105-1.pep		MPTVRFTESVSKQDLDALE	FAKASYGAES	CWKTLYLNL	PLGNLSPEW	VERVKKDWEAG	
orf105ng-1		MPTVRFTESVSKQDLDALE	FAKASYGAES	CWKTLYLNL	PLGNLSPEWA	ERIKDWEG	
		10	20	30	40	50	60
orf105-1.pep		CSESSDGIFLNADGWPMGG	RLQHLALGWH	CAGLLDGWRNE	CFDLTDGGG	NPLFTLERA	
orf105ng-1		CSESSDGIFLNADGWPMGG	RLQHLARTWN	KAGLLHGWRNE	CFDLTDGGG	NPLFTLERA	
		70	80	90	100	110	120
orf105-1.pep		FRPFGLLSRAVHLNGLTES	DGRWHFWIGRR	SPHKAVD	PNKLDNTA	AAGGVSGGEMP	SEAVC
orf105ng-1		FRPFGLLSRAVHLNGLVES	NGRWHFWIGRR	SPHKAVD	PGKLDNIAGGGV	SGGEMPSEAVC	
		130	140	150	160	170	180
orf105-1.pep		FRPFGLLSRAVHLNGLTES	DGRWHFWIGRR	SPHKAVD	PNKLDNTA	AAGGVSGGEMP	SEAVC
orf105ng-1		FRPFGLLSRAVHLNGLVES	NGRWHFWIGRR	SPHKAVD	PGKLDNIAGGGV	SGGEMPSEAVC	

-333-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf105-1.pep	RESSEEAGLDKTLPLIRPVSQHLRSVSRGVHNEILYVFDAVLPETFLPENQDGEVAG					
	orf105ng-1	RESSEEAGLDKTLFPLIRPVSRLHSLRPVSRGVHNEILYVFDAVLPETFLPENQDGEVAG					
		190	200	210	220	230	240
		250	260	270	280	290	
	orf105-1.pep	FEKMDIGGLLDAMLSGNMMHDAQLVTLDACRYGLIDAAHPLSEWLDGIRLX					
10	orf105ng-1	FEKMDIGGLLDAMLSKNMMHDAQLVTLDIFYRYGLIDAAHPLSEWLDGIRLX					
		250	260	270	280	290	

Furthermore, ORF105ng-1 (SEQ ID NO: 418) shows homology with a yeast enzyme (SEQ ID NO: 1139):

15 sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
)gi|1076928|pir||S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
 (Schizosaccharomyces pombe))gi|666111 (X84417) thiamin pyrophosphokinase
 [Schizosaccharomyces pombe])gi|2330852|gnl|PID|e334056 (Z98533) thiamin
 20 pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
 Score = 105 bits (259), Expect = 4e-22
 Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

Query: 268 NKAGLLHGWNRNECFDLTDGGGNPLFTLERAAPFRPFGLLSRAVHLNGLVESNGRW--HFWI 441
 N G+ WRNE + + P+ +ER F FG LS VH + + W+
 Sbjet: 96 NTFGIADQWRNELYTVYGKSKKPVLAVERRGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155

25 Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESSEEAGLDKTLFPLIRPVSRLHSLR 621
 RRSP K P LDN GG++ G+ + +E SEEA LD + LI P + ++
 Sbjet: 156 PRRSPTKQTWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNLI-PCGTVSYIK 214

Query: 622 PVSRG-VHNEILYVFDAVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVT 798
 R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
 30 Sbjet: 215 MEKRHWIQPELQYVFDLPVDDLVIPRINDGEVAGFSLPLNQLVHELELKSFKPNCALVL 274

Query: 799 LDAFYRYGLIDAAHP 843
 LD R+G+I HP
 Sbjet: 275 LDFLIRHGIITPQHP 289

35 Based on this analysis, including the presence of a putative transmembrane domain in the
 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID
 40 419>] (SEQ ID NO: 419):

1 ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCC

-334-

51 CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
 101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
 151 TTGATATTTG GTAACATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
 201 ACCTGCATCG GCGTAATCA GGGTGTATGC ACCGgATACG rGkACAATTA
 5 251 CAGCGAAATT CGTGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
 301 TTTGCGCTTT CGACCTCACG TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
 351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAACTGG
 401 GTCGTCTGAA GCTGATACAC GGAATGAAA CGCGCAGCcT TAAAGCAACT
 451 GTCGAACGTT TGGAAAACCA GGAATCCAT ATTCGCAAC AGATAGACGG
 10 501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
 551 TCCTATCCGC .CAATGA

This corresponds to the amino acid sequence [<SEQ ID 420; ORF107>] (SEQ ID NO: 420; ORF107):

15 1 MNRPKQPPFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
 51 LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
 101 FALSTSRFGA GGSVQQQLKT EAVLKKTAE QELGRLKLIH GNETRSLKAT
 151 VERLENQELH ISQQIDGQKR RIRLAEEMLO KYRFLSXQ*

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF107 (SEQ ID NO: 420) shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) (SEQ ID NO: 422) from strain A of *N. meningitidis*:

25	orf107.pep	10 20 30 40 50 60	MNRPKQPPFRPEVAVARQTSLTGKVILTRPLSFSLWTTTFASISALLIILFLIFGNYTRKT
	orf107a	10 20 30 40 50 60	MNRPKQPPFRPEVAVARQTSLTGKVILTRPLSFSLWTTTFASISALLIILFLIFGNYTRKT
30	orf107.pep	70 80 90 100 110 120	TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT
	orf107a	70 80 90 100 110 120	TVEGQILPASGVIRVYAPDTGTITAKFXEDGEKVKAGDKLFALSTSRFGAGDSVQQQLKT
35	orf107.pep	130 140 150 160 170 180	EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO
	orf107a	130 140 150 160 170 180	EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO
40	orf107.pep	189	KYRFLSXQX
	orf107a	190 200 210 220 230	KYRFLSANDAVPKQEMMNVKAEELLEQKAKLDAYRREEVGLLQEIRTQNLTLXSLPQAAX

45 The complete length ORF107a nucleotide sequence [<SEQ ID 421>] (SEQ ID NO: 421) is:

Based on the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 50

- 5 The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 425>] (SEQ ID NO: 425):

```

1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
51  GTGCGGCAAA TCCGTAAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCATAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCGAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
15  401 GCAAACCTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCT CCGCCGCCAT TATTGA

```

- 20 This corresponds to the amino acid sequence [<SEQ ID 426; ORF108>] (SEQ ID NO: 426; ORF108):

```

1  MLNTFFAVLG GCLLLXPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVG NIA EDGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

- 25 Further work revealed the following DNA sequence [<SEQ ID 427>] (SEQ ID NO: 427):

```

1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCGC
51  CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA
30  101 GCGCGCCGAA ACCGGTTTTC AAAGTCAAAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCGAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
35  401 GCAAACCTGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCT CCGCCGCCAT TATTGA

```

- 40 This corresponds to the amino acid sequence [<SEQ ID 428; ORF108-1>] (SEQ ID NO: 428; ORF108-1):

```

1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVG NIA EDGKLT DYL VSHAALQPYQ

```


151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. gonorrhoeae*

- 5 ORF108 (SEQ ID NO: 426) shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) (SEQ ID NO: 430) from *N. gonorrhoeae*:

	orf108.pep	MLNTFFAVLGGCLLXLPCGKSVENTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE	60
	orf108ng	MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALGQSSE	60
10	orf108.pep	GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT	120
	orf108ng	GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
	orf108.pep	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181
15	orf108ng	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181

ORF108-1 (SEQ ID NO: 428) shows 92.3% identity with ORF108ng (SEQ ID NO: 430) over the same 181 aa overlap:

	orf108-1.pep	MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE	60
20	orf108ng-1	MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALGQSSE	60
	orf108-1.pep	GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT	120
	orf108ng-1	GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120
	orf108-1.pep	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181
25	orf108ng-1	LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH	181

The complete length ORF108ng nucleotide sequence [<SEQ ID 429>] (SEQ ID NO: 429) is:

30	1	ATGCTGAAAa	tacctTTTGC	CGTGTtgggc	ggCtgccctGC	TGCTTGCCGC
	51	CTGCGGCAAA	TCCGAAAATa	cggcggaACA	GCCGCAAAAT	gcggCACAAA
	101	GCGCGCCGAA	ACCGGTTTTC	AAAGTCAAAT	ACATCGACAA	TACGGCGATT
	151	GCCGGTTTGG	CTTTGGGACA	AAGTAGCGAA	GGCAAAACCA	acgacgGCAA
	201	AAAACAAATC	AGTTATccgA	TTAAAGGCTT	GCCGGAACAA	Aacgccgtcc
	251	gGCTGACCGG	AAAGCATCCC	AACGACTTGG	AagccgtcgT	CGGCAAAATGT
35	301	ATGGAAACCG	ACGGAAAGGA	CGCGCCTTCG	GGCTGGGCGG	AAAACGGCGT
	351	GTGCCATACC	TTGTTTGCCA	AACTGGTGGG	CAATATCGCC	GAAGACGGCG
	401	GCAAACCTGAC	TGATTACCTG	ATTTGCGATT	CCGCCCTGCA	ACCCTATCAG
	451	GCAGGCAAAA	GCGGTATGTC	CGCCGTGCAG	AACGACGCT	ATGTGCTGGA
40	501	AATCGACAGC	GagggGGCGT	TTTATttccg	cgcgcgcat	tattgA

This encodes a protein having amino acid sequence [<SEQ ID 430>] (SEQ ID NO: 430):

```

1  MLKIPFAVLG GCLLLAACGK SENTAEQPQN AAQSAPKPVF KVKYIDNTAI
51  AGLALQSSSE GKTNDGKKQI SYPIKGLPEQ NAVRLTGKHP NDLEAVVGKC
101 METDGDAPS GWAENGVCHT LFAKLVGNIA EDGGKLTDL IHSALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 51

The following DNA sequence was identified in *N.meningitidis* [<SEQ ID 431>] (SEQ ID NO: 431):

```

1  ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
51  CGgATTtATC GATgcatTg cGggCGGGGG TGGTTTGATT ACGCTGCCCCG
101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
151 CTGCAAgCAG CCGCTGCTAC GTTTTcAGCT ACGGTTTCTT TTGCACGCAA
201 AGGTTTgATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
251 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301 CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTcGCAC TGTATTTTGT
351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTTCTGTT cGGGCTGACG GTCGC .ACCG .CTTTTGGGTT TTTACGACGG
451 TGTGTTTCGA CCGGTGTCG GCTCGTTTTT TCTGATTGCC TTTATGTTT
501 TGCTCGGCTG CAAgCTGTTG AACGCGATGT CTTACACCAA ATTGGCGAAC
551 GTTGCCtGCA ATCTTGGTTC GCTATCGGTA TTCCTGCTGC ACGGTTcGAT
601 TATTTTCCCG ATTGCGGCAA CGatGGCGGT CGGTGCGTTT GTCGgtGCGA
651 ATTTAgTGc GAGATTtGCC GTaCgtTCG GTTCGAAGCT GATTAA

```

This corresponds to the amino acid sequence [<SEQ ID 432; ORF109>] (SEQ ID NO: 432; ORF109):

```

1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VXTAFGFLRR
151 CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD
201 YFPDCGNDGG RCVCRCEFR C EICRTLRFEA D*

```

Further work revealed the following DNA sequence [<SEQ ID 433>] (SEQ ID NO: 433):

```

1  ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
51  CGGATTtATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCCCG
101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
151 CTGCAAGCAG CCGCTGCTAC GTTTTcAGCT ACGGTTTCTT TTGCACGCAA
201 AGGTTTgATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
251 TAGGCGGCGT GGCCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTcGCAC TGTATTTTGT

```

351	TTTTTCGCCC	AAGCTCGACG	GCAGTAAGGA	AGGCAAAGCC	AGAATGTCTT
401	TTTTTCTGTT	CGGGCTGACG	GTCGCACCGC	TTTTGGGTTT	TTACGACGGT
451	GTGTTCCGAC	CGGGTGTCCG	CTCGTTTTTT	CTGATTGCCT	TTATTGTTTT
501	GCTCGGCTGC	AAGCTGTTGA	ACGCATGTC	TTACACCAAA	TTGCGGAACG
551	TTGCCTGCAA	TCTTGGTTCG	CTATCGGTAT	TCCTGCTGCA	CGGTTTCGATT
601	ATTTTCCCCG	TTGCGGCAAC	GATGGCGGTC	GGTGCCTTTG	TCGGTGCGAA
651	TTTAGTGTGC	AGATTTGCCG	TCCGCTTCGG	TTCCAAGCTG	ATTAAGCCCG
701	TGCTGATTGT	CATCAGCATT	TGCATGGCTG	TGAAATTGTT	GATAGACGAG
751	AGAAATCCGC	TGTATCAGAT	GATTGTTTCG	ATGTTTTTAA	

This corresponds to the amino acid sequence [[<SEQ ID 434; ORF109-1>](#)] ([SEQ ID NO: 434; ORF109-1](#));

1	MEDLYIILAL	GLVAMIAGFI	DAIAGGGGLI	TLPALLLAGI	PPVSAIATNK
51	LQAAAATFSA	TVSFARKGLI	DWKKGLPIAA	ASFVGGVAGA	<u>LSVSLVSKDI</u>
101	LLAVVPVLLI	FVALYVFVSP	KLDGSKEGKA	RMSFFLFGLT	VAPLLGFPYDG
151	VFGPGVGSFF	<u>LIAFIVLLGC</u>	KLLNAMSYTK	LANVACNLGS	LSVFL LHGSI
201	IFPIAATMAV	GAFVGANLGA	RFAVRFGSKL	IKPLLIVISI	SMAVKLLIDE
251	RNPLYOMIVS	MF*			

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 (SEQ ID NO: 432) shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) (SEQ ID NO: 436) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf109.pep	MEDLYIILALGLVAMIAGFDIAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAAATFSA					
orf109a						
	10	20	30	40	50	60
	MEDLYIILALGLVAMIAGFDIAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAAATFSA					
	70	80	90	100	110	120
orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
orf109a						
	70	80	90	100	110	120
	TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
	130	140	150	160	170	180
orf109.pep	KLDGSKEGKARMSFFLFGTLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ					
orf109a						
	130	140	150	160	170	180
	KLDGSKEGKARMSFFLFGTLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK					

The complete length ORF109a nucleotide sequence [<SEQ ID 435>] (SEQ ID NO: 435) is:

1	ATGGAAGATT	TATACATAAT	ACTCGCTTTG	GTTTGGTTG	CGATGATTGC
51	CGGATTTATC	GATGCGATTG	CGGGTGGGGG	TGGTTTGATT	ACGTCGCCTG
101	CACTCTTGTT	GGCAGGTATT	CCTCCCGTGT	CGGCAATTGC	CACCAACAAG
151	CTGCAAGCAG	CCGCTGCTAC	GTTTTCGGCT	ACGGTTTCTT	TTGCACGC AA
201	AGGTTTGATT	GATTGGAAGA	AAGGTCCTCC	GATTGCGCGA	GTGCTGTTTG
251	CAGGCGCGCT	GGTCGGTGCA	TTATCGGTCA	GCTTGGTTTC	CAAAGATATT

-340-

5
10

```

301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTCTGTTT CGGTCTGACG GTTGCAACCAC TTTTGGGTTT TTACGACGGT
451 GTGTTTCGGAC CGGTGTGCGG CTCGTTTTTTT CTGATTGCCT TTATTGTTTT
501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
551 TTGCCTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTCGATT
601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
651 TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

```

This encodes a protein having amino acid sequence [SEQ ID 436] (SEQ ID NO: 436):

15

```

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151 VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
251 RNPLYQMIVS MF*

```

20 ORF109a (SEQ ID NO: 436) and ORF109-1 (SEQ ID NO: 434) show 99.2% identity in 262 aa overlap:

25

```

           10      20      30      40      50      60
orf109a.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
orf109-1    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
           10      20      30      40      50      60

           70      80      90     100     110     120
orf109a.pep TVSFARKGLIDWKKGLPIAAASFAGGVVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
orf109-1    TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
           70      80      90     100     110     120

           130     140     150     160     170     180
orf109a.pep KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFPGVGSFFLIAFIVLLGCKLLNAMS YTK
orf109-1    KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFPGVGSFFLIAFIVLLGCKLLNAMS YTK
           130     140     150     160     170     180

           190     200     210     220     230     240
orf109a.pep LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
orf109-1    LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
           190     200     210     220     230     240

           250     260
orf109a.pep SMAVKLLIDERNPLYQMIVSMFX
orf109-1    SMAVKLLIDERNPLYQMIVSMFX
           250     260

```

45

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 (SEQ ID NO: 432) shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) (SEQ ID NO: 438) from *N. gonorrhoeae*:

```

5      orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60
      orf109ng    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60

      orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120
      orf109ng    TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120

10     orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180
      orf109ng    KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180

      orf109.pep  IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRTLRFEAD  231
      orf109ng    IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRPLRFEAD  231
15

```

An ORF109ng nucleotide sequence [<SEQ ID 437>] (SEQ ID NO: 437) was predicted to encode a protein having amino acid sequence [<SEQ ID 438>] (SEQ ID NO: 438):

```

20     1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
      51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
      101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VATAFGFLRR
      151 CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD
      201 YFPDCGNDGG RCVCRCEFR C EICRPLRFEA D*

```

Further work revealed the following gonococcal DNA sequence [<SEQ ID 439>] (SEQ ID NO: 439):

```

30     1  ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTTGGTTG CGATGATCGC
      51  CGGATTTATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCTG
      101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
      151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
      201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
      251 CAGGCGGCGT GGTGCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
      301 TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
      351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
      401 TTTTCTTATT CGGGCTGACG GTTGCACCGC TTTTGGGTTT TTACGACGGT
      35  451 GTGTTTCGAC CGGGTGTGCG CTCGTTTFTT CTGATTGCCT TTATGTGTTT
      501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG
      551 TTGCTTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
      601 ATTTTCCCGA TTGTGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
      651 TTTAGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
      40  701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
      751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

```

This corresponds to the amino acid sequence [<SEQ ID 440; ORF109ng-1>] (SEQ ID NO: 440; ORF109ng-1):

```

45     1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK

```

51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
 151 VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
 201 IFPIVATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
 251 RNPLYQMIVS MF*

ORF109ng-1 (SEQ ID NO: 440) and ORF109-1 (SEQ ID NO: 434) show 98.9% identity in 262 aa overlap:

```

10      orf109ng-1.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      orf109-1      MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      10          20          30          40          50          60

15      orf109ng-1.pep TVSFARKGLIDWKKGLPIAAASFAGGVVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      orf109-1      TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70          80          90          100         110         120

20      orf109ng-1.pep KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      orf109-1      KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      130         140         150         160         170         180

25      orf109ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      orf109-1      LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      190         200         210         220         230         240

30      orf109ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
      orf109-1      SMAVKLLIDERNPLYQMIVSMFX
      250         260

```

35 In addition, ORF109ng-1 (SEQ ID NO: 440) shows homology to a hypothetical *Pseudomonas* protein (SEQ ID NO: 1140):

```

40      sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
      )gi|94984|pir|I38164 hypothetical protein 9 - Pseudomonas sp )gi|551929 (M62866)
      ORF9 [Pseudomonas denitrificans] Length = 261
      Score = 175 bits (439), Expect = 3e-43
      Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

      Query: 41 PPVSAIATNKLQXXXXXXXXXXXXXXXXXKGLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKDI 100
      PP+ + TNKLQ R+G ++ K+ LP+ D+
      Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

45      Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFF 160
      L A++P LLI +ALYF P + G + +R++ F+F LT+ PL+GFYDGVFGPG GSFF
      Sbjct: 103 LKAILPFLLI AIALYFGLKPNM-GDVQHSRVTPFVFTLTLPVPLIGFYDGVFGPGTGSFF 161

```

Query: 161 LIAFIVLLGCKLLNAMS YTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGA 220
 ++ F+ L G +L A ++TK N N+G+ VFL G++++ + M +G F+GA +G+
 Sbjct: 162 MLGFVTLAGFGVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLGAQVGS 221

Query: 221 RFAVRFGSKLIKPLLLIVISISMAVKLLIDERNPL 254
 R+A+ G+K+IKPLL+++SI++A++LL D +PL
 Sbjct: 222 RYAMAKGAKIIKPLLIVISIALAIRLLADPTHPL 255

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 441>] (SEQ ID NO: 441):

```

15      1  ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCAG
      51  CCGCAGACGG ATTATTGGT CAAATTCGGA TCGTTTGGG CGAG.ATTTT
      101  TGGTTTCTG GGACTGTATG ACGTCTATGC TTCGCGATGG TTGTGCGTTA
      151  TCATGATGTT TTTGGTGGTT TCTACCAGTT TGTGCCTGAT TCGCAATGTG
      201  CCGCCGTTCT GCGCGGAAAT GAAGTCTTTT CGGAAAAGG TTAAAGAAAA
20      251  ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGATGTA AAAATTGCGC
      301  CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTCA GGGGAAAACC
      351  ATTAACCGTG AAGACGGGTC GGTTCGATT GCCGCCAAA AAGGCACAAT
      401  GAACAAATGG GGTATATCT TTGCCATGT TGCTTGATT GTCATTGACC
      451  TGGGCGGGT GATAGACAGT AACCTGCTGT TGAACTGGG TATGCTGACC
25      501  GGTGCGATTG TTCCGGACAA TCAGGCGGTT TATGCCAAGG ATTTC.AAGC
      551  CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTAGGG GCAACGTCAA
      601  TATTTCCG.A GGGGCAGaT GCGGATGTGG TTTTCCTGA

```

This corresponds to the amino acid sequence [<SEQ ID 442; ORF110>] (SEQ ID NO: 442; ORF110):

```

      1  ..LLGIASVIGT LLQONQPQTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI
      51  MMFLVVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP
      101  EVAKRYLEVQ GFQGTINRE DGSVLIAAKK GTMNKGYIF AHVALIVICL
      151  GGLIDSNLLL KLGMLTGRIF RTIRRFMPRI XKPESXFGCV QSLI*GQRQY
35      201  FXRGRVRMWF S*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 (SEQ ID NO: 442) shows 91.5% identity over a 188aa overlap with ORF88a (SEQ ID NO: 332) from strain A of *N. meningitidis*:

-344-

		10	20	30	40	50	60
	orf88a.pep	MSKSRRSPPLLSRPWF	FAFFSSMRFAVALLSLLGIASVIGTVLQQNQ	PQTDYLVKFGSFWA			
5	orf110			LLGIASVIGTLLQQNQ	PQTDYLVKFGSFWA		
				10	20	30	
		70	80	90	100	110	120
	orf88a.pep	QIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNV	PPFWREMKSFREKVKEKSLAAMRH			
10	orf110	XIFGFLGLYDVYASAWFV	VIMMFLVVSTSLCLIRNV	PPFWREMKSFREKVKEKSLAAMRH			
		40	50	60	70	80	90
		130	140	150	160	170	180
	orf88a.pep	SSLLDVKIAPEVAKRYLEVQ	GFGKTINREDGSVLIAAK	KGTMNKGWYIFAHVALIVICL			
15	orf110	SSLLDVKIAPEVAKRYLEVQ	GFGKTINREDGSVLIAAK	KGTMNKGWYIFAHVALIVICL			
		100	110	120	130	140	150
		190	200	210	220	230	240
	orf88a.pep	GGLIDSNLLLKLGLMTGR	IVPDNQAVYAKDFKPESIL	GASNLSFRGNVNISEGQSADVVF			
20	orf110	GGLIDSNLLLKLGLMTGR	IFRTIRRFMPRIXKPESX	FGCVQSLIXGQRQYFXRGRV	RMWF		
		160	170	180	190	200	210
		250	260	270	280	290	300
	orf88a.pep	LNADNGILVQDLPFEVKL	KKFHIDFYNTGMPRDFAS	DIETDKATGEKLERTIRV	NHPLT		
25	orf110	SX					

However, ORF88 (SEQ ID NO: 328) and ORF110 (SEQ ID NO: 442) do not align, because they represent two different fragments of the same protein.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 (SEQ ID NO: 442) shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) (SEQ ID NO: 444) from *N. gonorrhoeae*:

	orf110.pep	LLGIASVIGTLLQQNQPQTDYLVKFGSFWA	30
	orf110ng	MSKSRI SPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPQTDYLVKFGPFWT	60
35	orf110.pep	XIFGFLGLYDVYASAWFVVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	90
	orf110ng	RIFDFLGLYDVYASAWFVVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQFGQGKTINREDGSVLIAAKKGTMNKGWYIFAHVALIVICL	150
	orf110ng	SSLLDVKIAPEVAKRYLEVRFQFGKTVSREDGSVLIAAKKGTMNKGWYIXAHVALIVICL	180
40	orf110.pep	GGLIDSNLLLKLGLMTGRIFRTIRRFMPRIXKPESXFGCVQSLIXGQRQYFXRGRVRMWF	210
	orf110ng	GRLINXNLLLKLGLMAGSIFRNNRRVMPRISKPESIWGGVQSLIKGQRQYFQRGKVRMWF	240


```

orf110.pep      S   211
                |
orf110ng        S   241

```

- 5 The complete length ORF110ng nucleotide sequence [<SEQ ID 443>] (SEQ ID NO: 443) is predicted to encode a protein having amino acid sequence [<SEQ ID 444>] (SEQ ID NO: 444):

```

10      1  MSKSRIPTL  LSRPWFAFFS  SMRFAVALLS  LLGIASVIGT  VLOQNQPQTD
      51  YLVKFGPFWT  RIFDFLGLYD  VYASAWFVVI  MMFLVVSTSL  CLIRNVPPFW
     101  REMKSFREKV  KEKSLAAMRH  SSLLDVKIAP  EVAKRYLEVVR  GFQGKTVSRE
     151  DGSVLIAAKK  GTMNKWDYIX  AHVALIVICL  GRLINXNLLL  KLGMLAGSIF
     201  RNNRRVMPRI  SKPESIWGGV  QSLIKGQRQY  FQRGKVRMWF  S*

```

Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 53

The following DNA sequence was identified in *N.meningitidis* [<SEQ ID 445>] (SEQ ID NO: 445):

```

20      1  ATGCCGTCTG  AAACACGCCT  GCCGAAC TTT  ATCCGCGTCT  TGATATTTCG
      51  CCTGGGTTTC  ATCTTCCTGA  ACGCCTGTTC  GGAACAAACC  GCGCAAACCG
     101  TTACCCTGCA  AGGCGAAACG  ATGGGCACGA  CCTATACCGT  CAAATACCTT
     151  TCAAATAATC  GGGACAAACT  CCCCTCACCT  GCCGAAATAC  AAAAACGCAT
     201  CGATGACGCG  CTTAAAGAAG  TCAACCGGCA  GATGTCCACC  TATCAGCCCG
     251  ACTCCGAAAT  CAGCCGGTTC  AACCAACACA  CAGCCGGCAA  GCCCCCTCCG
     25  301  ATTTCAAGCG  ACTTCGCACA  CGTTACTGCC  GAAGCCGTCC  GCCTGAACCG
     351  CCTGACACAC  GGCGCGCTGG  ACGTAACCGT  CGGCCCTTG  GTCAACCTTT
     401  GGGGATTTCG  CCCCACAAA  TCCGTTACCC  GTGAACCGTC  GCCGGAACAA
     451  ATCAAACAGG  CGGCATCTTA  TACGGGCATA  GACAAAATCA  TTTTGAAACA
     501  AGGCAAAGAT  TACGTTCCCT  TGAGCAAAAC  CCACCCCAAG  GCCTATTTCG
     30  551  ATTTATCTTC  GATTGCCAAA  GGCTTCGGCG  TTGATAAAGT  TGCGGGCGAA
     601  CTGGAAAAAT  ACGGCATTCA  AAATTATCTG  GTCGAAATCG  GCGGCGAGTT
     651  GCACGGCAAA  GGCAAAACG  CGCGCGGCGA  ACCGTGGCGC  ATCGGTATCG
     701  AGCAGCCCAA  TATCGTCCAA  GGCGGCAATA  CGCAGATTAT  CGTCCCCTCG
     751  AACAAACGTT  CGCTTGCCAC  TTCCGGCGAT  TACCGTATTT  TCCACGTCGA
     35  801  TAAAAACGGC  AAACGCCTCT  CCCATATCAT  CAACCCGAAC  AACAAACGAC
     851  CCATCAGCCA  CAACCTCGCC  TCCATCAGCG  TGGTCGCAGA  CAGTGCGATG
     901  ACGGCGGACG  GCTTGTCCAC  AGGATTATTC  GTATTGGGCG  AAACCGAAGC
     951  CTAAAGCTG  GCAGAGCGCG  AAAAATCGC  TGTTTTCCTG  ATTGTCAGGG
     1001  ATAAAGGCGG  CTACCGCACC  GCCATGTCTT  CCGAATTGTA  AAACTGCTC
     40  1051  CGCTAA

```

This corresponds to the amino acid sequence [<SEQ ID 446; ORF111>] (SEQ ID NO: 446; ORF111):

```

45      1  MPSETRLPNF  IRVLIFALGF  IFLNACSEQT  AQTVTLQGET  MGTITYTVKYL
     51  SNNRDKLPSP  AEIQKRIDDA  LKEVNRQMST  YQPDSEISRF  NQHTAGKPLR

```

-346-

5
101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFGPDK SVTREPSPEQ
151 IKQAASYTGI DKIIILKQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE
201 LEKYGIQNYL VEIGGELHGK GKNARGEPR IGIEQPNIVQ GGNTQIIIVPL
251 NNRLATSGD YRIFHVDKNG KRLSHIINPN NKRPIHNLASISVXADSAM
301 TADGLSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
351 R*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 (SEQ ID NO: 446) shows 96.9% identity over a 351aa overlap with an ORF (ORF111a)

10 (SEQ ID NO: 448) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf111a.pep	MPSETRLPNF	IRTLIFALS	FIFLNACSE	QTAQTVTLQ	GETMGTTYT	VKYLSNNRDXLPSP
	orf111	MPSETRLPNF	IRVLIFALG	FIFLNACSE	QTAQTVTLQ	GETMGTTYT	VKYLSNNRDKLPSP
15		10	20	30	40	50	60
	orf111a.pep	70	80	90	100	110	120
	orf111	AEIQXRIDDALKEVN	QRMSTYQPDSEIS	RFNQHTAGKPLR	ISSDFAHVTA	EAVHLNRLTH	
20		70	80	90	100	110	120
	orf111a.pep	130	140	150	160	170	180
	orf111	GALDVTVGPLVNLWG	FPGDKSVTREPS	PEQIKQAASYT	GIDKIIILKQ	GKDYASLSKTHPK	
25		130	140	150	160	170	180
	orf111a.pep	190	200	210	220	230	240
	orf111	AYLDLSSIAKGF	GVDXVAGELEKY	GIQNYLVEIGG	ELHGKXKNAR	GEPWRIGIEQ	PNIVQ
30		190	200	210	220	230	240
	orf111a.pep	250	260	270	280	290	300
	orf111	GGNTQIIIVPLNN	RSXATSGDYRIF	HVDKSGKRLSHI	INPNKRPIHNL	ASISVXADSAM	
35		250	260	270	280	290	300
	orf111a.pep	310	320	330	340	350	
	orf111	TADGXSTGLFVL	GETEALKLAERE	KLAVFLIVRD	KGGYRTAMSSE	FEKLLRX	
40		310	320	330	340	350	

The complete length ORF111a nucleotide sequence [<SEQ ID 447>] (SEQ ID NO: 447) is:

45
1 ATGCCGTCCTG AAACACGCCT GCCGAACCTT ATCCGCACCT TGATATTTGC
51 CCTGAGTTT ATCTTCTGA ACGCTGTTC GGAACAAACC GCGCAAACCG
101 TTACCCTGCA AGGTGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT

-347-

5
10
15
20

```

151 TCAAATAATC GGGACNAACT CCCNTCACCT GCCGAAATAC AAAANCGCAT
201 CGATGACGCG CTAAAGAAG TCAACCGGCA GATGTCCACC TATCAGCCCG
251 ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCTCCGC
301 ATTTCAAGCG ACTTCGCACA CGTTACTGCC GAAGCCGTCC ACCTGAACCG
351 CCTGACACAC GGC GCGCTGG ACGTAACCGT CGGCCCTTG GTCAACCTTT
401 GGGGATTCCG CCCCACAAA TCCGTACCC GTGAACCGTC GCCGGAACAA
451 ATCAAACAAG CAGCATCTTA TACGGGCATA GACAAAATCA TTTTGAAACA
501 AGGCAAAGAT TACGCTTCCT TGAGCAAAAC CCACCCCAAG GCCTATTTGG
551 ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATNANGT TGCGGGCGAA
601 CTGGA AAAAT ACGGCATCA AAATTATCTG GTCGAAATCG GCGGNGAGTT
651 GCACGGCAAA GNCAAAAACG CGCGCGGCGA ACCTTGGCGC ATCGGCATCG
701 AACAGCCCAA CATCGTCCAA GCGGCAATA CGCAGATTAT CGTCCCGCTG
751 AACACCGTT CGNTTGCCAC TTCCGGCGAT TACCGTATT TCCACGTCGA
801 TAAAAGCGG AAACGCCTCT CCCATATCAT TAATCCGAAC AACAAACGAC
851 CCATCAGCCA CAACCTCGCC TCCATCAGCG TGNTCGCAGA CAGTGCATG
901 ACGGCGGACG GCTTNTCCAC AGGATTATTC GTATTGGGCG AAACCGAAGC
951 CTTAAAGCTG GCAGAGCGCG AAAAACTCGC TGTTTTCCTG ATTGTCAGGG
1001 ATAAAGGCGG CTACCGCACC GCCATGTCTT CCGAATTGA AAAACTGCTC
1051 CGCTAA

```

This encodes a protein having amino acid sequence [SEQ ID 448] (SEQ ID NO: 448):

25
30

```

1 MPSETRLPNF IRTLIFALSF IFLNACSEQT AQTVTLQGET MGTITYTVKYL
51 SNNRDXLPSP AEIQXRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR
101 ISSDFAHVTA EAVHLNRLTH GALDVTVGPL VNLWGFGPDK SVTREPSPEQ
151 IKQAASYTGI DKIIILKQKD YASLSKTHPK AYLDLSSIAK GFGVDXVAGE
201 LEKYGIQNYL VEIGGELHGK XKNARGEPRW IGIEQPNIVQ GGNTQIIIVPL
251 NNRSXATSGD YRIFHVDKSG KRLSHIINPN NKRPISHNLA SISVXADSAM
301 TADGXSTGLF VLGETEALKL AEREKLAFLV IVRDKGGYRT AMSSEFEKLL
351 R*

```

30 Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 (SEQ ID NO: 446) shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) (SEQ ID NO: 450) from *N. gonorrhoeae*:

35
40
45

```

          10      20      30      40      50      60
orf111ng  MPSETRLPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGTITYTVKYLNNRDKLPSP
          |||||:|:|||||
orf111    MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTITYTVKYLNNRDKLPSP
          10      20      30      40      50      60

          70      80      90      100     110     120
orf111    AKIQKRIDDALKEVNRQMSTYQTDSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH
          |:|||||
orf111    AEIQKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH
          70      80      90      100     110     120

          130     140     150     160     170     180
orf111ng  GALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGIDKIIILQQGDYASLSKTHPK
          |||||:|:|||||
orf111    GALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGIDKIIILKQGDYASLSKTHPK
          130     140     150     160     170     180

          190     200     210     220     230     240
orf111ng  AYLDLSSIAKGFGVDKVAGELEKYGIQNYLVEIGGELHGKGKNAHGEPWRIGIEQPNIIQ

```

[illegible]

15 The complete length ORF111ng nucleotide sequence [<SEQ ID 449>] (SEQ ID NO: 449) is:

	1	ATGCCGCTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGCCT	TGATATTTGC
	51	CCTGGGTTC	ATCTTCCTGA	ACGCCTGTTC	GGaaccaaacC	GCGCAaaccg
	101	TTACCCTGCA	AGGCGAAAacg	aTGGGTACGA	CCTATACCGT	CAAATACCTT
20	151	TCAAATAATC	GGGACAAACT	CCCCTCCCCT	GCCAAAAATAC	AAAAGCGCAT
	201	TGATGATGCG	CTTAAAGAAG	TCAACCGGCA	GATTGTCCACC	TACCAGACCG
	251	ATTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCTCCGC
	301	ATTTCAAGCG	ATTTTCGCACA	CGTTACCGCC	GAAGCCGTCC	GCCTGAACCG
	351	CCTGACTCAC	GGCGCACTGG	ACGTAACCGT	CGGCCCTTTG	GTCAACCTTT
25	401	GGGGGTTTCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAAACAA
	451	ATCAAAACAG	CGGCATCTTA	TACGGGCATA	GACAAAAATCA	TTTTGCAACA
	501	AGGCAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAA	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATAAAGT	TGCGGGCGAA
	601	CTGGAAAAAT	ACGGCATTCA	AAATTATCTG	GTGCAAAatcg	gcgpcGAGTT
30	651	GCACGGCAAA	GGCAAAAAATG	CGCACGGCGA	ACCGTGGGCG	ATCGGTATAG
	701	AGCAACCCAA	TATCATCCAA	GgcgGCAata	CGCAGATTAt	cgtcccgctg
	751	aaCaaccgtt	cgctTGCCAC	TTCCGGCGAT	TAccgtaTTT	tccaagtcga
	801	TAAAAAcggc	aaacgccttt	cccacaTCAT	CAATCCCaAC	aacAAACgac
	851	ccATCAGcca	caacctcgcc	tccatcagcg	tggtctcAGA	CAGTGCAATG
35	901	ACGGCGGACG	GTTTatCCAC	AGGATTATTT	GTTTTAGGCG	AAACCGAAGC
	951	CTTAAGGCTG	GCAGAACAAAG	AAAAACTCGC	TGTTTTCTTA	ATTGTCCGGG
	1001	ATAAGGACGG	CTACCGCACC	GCCATGTCTT	CCGAATTTGC	CAAGCTGCTC
	1051	CGCTAA				

This encodes a protein having amino acid sequence [SEQ ID 450] (SEQ ID NO: 450):

40	1	MPSETRLPNL	IRALIFALGF	IFLNACSEQT	AQTVTLQGET	MGTITYTVKYL
	51	SNNRDKLPSP	AKIQKRIDDA	LKEVNRQMS	YQTDSEISRF	NQHTAGKPLR
	101	ISSDFAHVTA	EAVRLNRLTH	GALDVTVGPL	VNLWGFPGDK	SVTREPSPEQ
	151	IKQAASYTGI	DKIILQQGKD	YASLSKTHPK	AYLDLSSIAK	GFGVDKVAGE
	201	LEKYGIQNYL	VEIGGELHGK	GKNAHGEPWR	IGIEQPNI IQ	GGNTQII VPL
45	251	NNRSLATSGD	YRIFHVDKNG	KRLSHIINPN	NKRPISHNLA	SISVVSDSAM
	301	TADGLSTGLF	VLGETEALRL	AEQEKLAVFL	IVRDKDGYRT	AMSSEFAKLL
	351	R*				

This protein shows homology with a hypothetical lipoprotein precursor (SEQ ID NO: 1141) from *H. influenzae*:

sp|P44550|YOJL_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR)gi|1074292|pir|4
 hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20))gi|1573128
 (U32702) hypothetical [Haemophilus influenzae] Length = 346
 Score = 353 bits (896), Expect = 9e-97
 Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)

Query: 7 LPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSPAKIQKR 66
 + LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +
 Sbjct: 1 MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTTYHVKYLDGSIATSE-KTHEE 58

Query: 67 IDDALKEVNRQMSTYQTDSEISRFNQHT-AGKPLRISSDFAHVTAEAVRLNRLTHGALDV 125
 I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV
 Sbjct: 59 IEAILKDVNAKMSTYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTGALDV 118

Query: 126 TVGPLVNLWGFPGDPKSVTREPSPEQIKQAASYTGIDKIIQQGKDYASLSKTHPKAYLDL 185
 TVGP+VNLWGFPG+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL
 Sbjct: 119 TVGPVNLWGFPEKRPEKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178

Query: 186 SSIAGFGVDKVGAGELEKYGIQNYLVEIGGELHGKGKNAHGEPWRIGIEQPNIIQGGNTQ 245
 SSIAGFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +

Sbjct: 179 SSIAGFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKPWQIAIEKPTTTGERAVE 238
 Query: 246 IIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLASISVVSDSAMTADGL 305
 ++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL
 Sbjct: 239 AVIGLNNMGMASGDYRIY-FEENGKRFHEIDPKTYPIQHHLASITVLAPTSMTADGL 297

Query: 306 STGLFVLGETEALRLAEQEKLAFLIVRDKDGRTAMSSEFAKL 349
 STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL
 Sbjct: 298 STGLFVLGEDKALEVAEKNNLAVYLIIRTDNGFVTKSSSAFKKL 341

- 25 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 54

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 451>] (SEQ ID NO: 451):

30 1 ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCGTCA
 51 AAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG
 101 GCGGCGCGGC TCGGACGGG TGGCGCAAAG GCGTGCAAAT CGGCGGCGAG
 151 GTGTTTGTAC GGCAAAATGA AGGCAGCCKa yTGGCAATCG GCGTGATGGG
 201 CGGCAGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG
 35 gCAGTGATTT GTATGTTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG
 301 TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA
 351 ACGTTTCAAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA
 401 CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTTGGC
 451 GAAGGCATTG TCGGAAAAGG CAATAATGTG CGGTTTACC TACAACCGCA
 40 GgCGCAGTTT ACCTACTTGG GCGTAAACGG CGGCTTTACC GACAGCGAGG
 551 GGACGGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC
 601 AtTCGGGCAA AAACCCGTTT TGCTTTGCGT AACGGTGTC ATCTTCAGCC
 651 TTTTGCCGCT TTTAATGTtT TGCACAGGTC AAAATCTTTC GGCGTGAAAA
 701 TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG
 45 751 TTCGGTATTG AAGCCGTTG GAAAGGCCAT ATGTCGCA..

This corresponds to the amino acid sequence [<SEQ ID 452; ORF35>] (SEQ ID NO: 452; ORF35):

```

5      1  ..PCRRQGDDVY AAHASRQKLW LRFIGGRSHQ NIRGGAAADG WRKGVQIGGE
      51  VFVRQNEGSX LAIGVMGGRA GQHASVNGKG GAAGSDLYGY GGGVYAAWHQ
     101  LRDKQTGAYL DGWLQYQRFK HRINDENRAE RYKTKGWTAS VEGGYNALVA
     151  EGIVGKGNNV RFYLQPPAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG
     201  IRAKTRFALR NGVNLQPFPA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR
     251  FGIEAGWKGH MSA..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF (SEQ ID NO: 452) and virg-h protein (SEQ ID NO: 1146) show 51% aa identity in 261aa overlap:

```

15      Orf35  5  QGDDVYAAHASRQKLWLRFIGGRSHQNIRGGAA-ADGWRKGVQIGGEVFVRQNEGSXLAI 63
          +  D++    R+  LWLR I G S+Q ++G  A  +G+RKGVQ+GGEVF  QNE + L+I
      virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQWVQKGTAPVEGYRKGVLGGEVFTWQNESNQLSI 455

      Orf35  64  GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDKQTGAYLDGWLQYQRFKH 121
          G+MGG+A Q ++ +          ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H
20      virg-h 456 GLMGGQAEQRSTFHNPDNDLTTGNVKGFGAGVYATWHQLQDKQTGAYADSWMQYQRFH 515

      Orf35  122 RINDENRAERYKTKGWTASVEGGYNALVAEGIVGKGNNVRFYLQPPAQFTYLGVNGGFTD 181
          RIN E+  ER+ +KG TAS+E GYNAL+AE  KGN++R YLQPPAQ TYLGVNG F+D
      virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTKKGNLSRVYLQPPAQLTYLGVNGKFS 575

      Orf35  182 SEGTAVGLLGSQWQSRAGIRAKTRFALRNGVNLQPFPAFNVLHRSKSFGEVMDGEKQTL 241
          SE  V LLGS Q Q+R G++AK +F+L  + ++PFAA N L+ +K FGVEMDGE++ +
25      virg-h 576 SENAHVNLGSRQLQTRVGVAQAQFSLYKNIAIEPFAAVNALYHNKPFGEVMDGERRVI 635

      Orf35  242 AGRTALEGRFGIEAGWKGHMS 262
          +TA+E + G+  K H++
      virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

```

30 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF35 (SEQ ID NO: 452) shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) (SEQ ID NO: 454) from strain A of *N. meningitidis*:

```

35      orf35.pep                                10      20      30
          PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIRG
          : ||||| ||||| ||||| ||||| |||||
      orf35a  QRLAIPEAEAVLYAQQAYAANTLFGRLRAADRGDDVYAADPSRQKLWLRFIGGRSHQNIRG
          310      320      330      340      350      360

          40      50      60      70      80      90

```

-351-

	orf35.pep	GAAADGWRKGVQIGGEVFVRQNEGSXLAIGVMGGRAGQHASVNGKGAAGSDLYGYGGGV					
	orf35a	GAAADGRRKGVQIGGEVFVRQNEGSRLAIGVMGGRAGQHASVNGKGAAGSYLHGYGGGV					
		370	380	390	400	410	420
5	orf35.pep	YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV					
	orf35a	YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGV					
		430	440	450	460	470	480
10	orf35.pep	GKGNNVRFYLQPQAQFTYLVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRNGVN					
	orf35a	GKGNNVRFYLQPQAQFTYLVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRNGVN					
		490	500	510	520	530	540
15	orf35.pep	LQPFAAFNVLHRSKSGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA					
	orf35a	LQPFAAFNVLHRSKSGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSARIGYKRTDGD					
		550	560	570	580	590	600
20	orf35a	KEAALSLKWLFX					
		610	620				

The complete length ORF35a nucleotide sequence [<SEQ ID 453>] (SEQ ID NO: 453) is:

25	1	ATGTT	CAGAG	CTCAG	CTTGG	TTCAA	AATACT	CGTTC	TACCA	AAATC	GGCGA
	51	CGATG	CCGAT	TTTTCA	TTTTT	CAGACA	AGCC	GAAAC	CCCGC	ACTTC	CCCATT
	101	ATTTT	TCCAG	CGGTAA	AAACC	GATCAA	AATT	CATCC	GAATA	TGGGT	TATGAC
	151	GAAAT	CAATA	TCCAAG	GTAA	AACTACA	AT	AGCGG	CATAC	TCGCC	GTGCA
	201	TAATAT	GCCC	GTTGTT	AAGA	AATATAT	TAC	AGATA	CTTAC	GGGGA	TATT
30	251	TAAAG	GATGC	GGTTAA	GAAG	CAATTAC	AGG	ATTATA	CAAA	AACAAG	ACCC
	301	GAAGCT	TGGG	AAGAAA	ATAA	AAAACGG	ACT	GAGGAG	GCGT	ATATA	GAAACA
	351	GCTTGG	ACCA	AAATTT	AGTA	TACTCAA	ACA	GAAAA	ACCC	GATTTA	ATTATA
	401	ATAAAT	TGGT	AGAAG	ATTCC	GTACTCA	CTC	CTCAT	AGTAA	TACAT	CACAG
	451	ACTAGT	CTCA	ACAAC	ATCTT	CAATAAAA		TTACAC	GTC	AAATC	GAAAA
35	501	CAAAT	CCCA	GTCG	CCGAC	AGGTGT	TGGA	ACTGA	CCAAG	ATGAC	GCTGA
	551	AAGATT	CCCT	TTGGGA	ACCG	CGCGCC	ATT	CCGAC	ATCCA	TATGT	GCGAA
	601	ACTTCC	GATA	ATGCC	GCAT	CCGCCT	GAA	ACGAA	AGATG	AAAAA	CTGAC
	651	CGTCCA	TAAA	GCGTAT	CAGG	GCGGTG	CGGA	TTTCT	CTGTT	GGCTA	CGACG
	701	TGCGGG	AGTC	GGACAA	ACCC	GCCCTG	ACCT	TTGA	AAGAAA	AGTC	AGCGGA
40	751	CAATCC	GGCG	TGGTT	TGGA	ACGCCG	GCCG	GAAAT	CTGA	AAACG	CTCGA
	801	CGGGCG	CAAA	CTGATT	GCGG	CGGAAA	AGG	AGACT	CTAAT	TCGTT	TGCGT
	851	TTAAAC	AAAA	TTACCG	CGAG	GGA	CTGTACG	AATTAT	TGCT	CAAGC	AATGC
	901	GAAGG	CGGAT	TTTGCT	TGGG	CGTGC	AGCGT	TTGGC	TATCC	CCGAG	GCGGA
	951	AGCGG	TTTTA	TATGCC	CAAC	AGGCTT	ATGC	GGCAA	AATACT	TTGTT	CGGGC
45	1001	TGCGT	GCCGC	CGACAG	GGG	GACGAC	GTGT	ATGCC	GCCGA	TCCGT	CCCGT
	1051	CAAAA	ATTGT	GGCTGC	GCTT	CATCGC	GCGC	CGTGC	CATC	AAAAT	ATACG
	1101	GGGCG	CGCG	GCTGCG	ACG	GGCGCG	CAA	AGGCG	TGCAA	ATCGG	CGGCG
	1151	AGGTG	TTTGT	ACGGCA	AAAT	GAAGG	CAGCC	GGCTG	GCAAT	CGGCG	TGATG
	1201	GGCGG	CAGGG	CTGGCC	AGCA	CGCAT	CAGTC	AACGG	CAAAG	GCGGT	GCGGC
50	1251	AGGCA	GTTAT	TTGCAT	TGGT	ATGGC	GGGG	TGTTAT	TGCT	GCGT	GGCATC
	1301	AGTTG	CGCGA	TAAACA	AAACG	GGTGC	GTATT	TGGAC	GGCTG	GTTGA	CAATAC
	1351	CAACG	TTTCA	AACACC	GCAT	CAATG	ATGAA	AACCG	TGCGG	AACG	TACAA
	1401	AACCA	AAGGT	TGGAC	GCCTT	CTGTC	GAAAG	CGGCT	TACAA	GCGCT	TGTGG
	1451	CGGAA	GGCGT	TGTCG	GAAAA	GGCAAT	AATG	TGCGG	TTTTA	CCTGA	ACCG
	1501	CAGGC	GCGAGT	TTACCT	ACTT	GGGCG	TAAAC	GGCGG	CTTTA	CCGAC	AGCGA

5

1551	GGGGACGGCG	GTCGGA CTGC	TCGGCAGCGG	TCAGTGGCAA	AGCCGCGCCG
1601	GCATTCGGGC	AAAAACCCGT	TTTGCTTTGC	GTAACGGTGT	CAATCTTCAG
1651	CCTTT TGCCG	CTTTTAATGT	TTTGCACAGG	TCAAAATCTT	TCGGCGTGGA
1701	AATGGACGGC	GAAAAACAGA	CGCTGGCAGG	CAGGACGGCG	CTCGAAGGGC
1751	GGTTCCGGCAT	TGAAGCCGGT	TGGAAGGCC	ATATGTCCGC	ACGCATCGGA
1801	TACGGCAAAA	GGACGGACGG	CGACAAAGAA	GCCGCATTGT	CGCTCAAATG
1851	GCTGTTTTGA				

This encodes a protein having amino acid sequence [<SEQ ID 454>] (SEQ ID NO: 454):

10	1	MFRAQLGSNT	RSTKIGDDAD	FSFSDKPKPG	TSHYFSSGKT	DQNSSEYGYD
	51	EINIQGKNYN	SGILAVDNMP	VVKKYITDTY	GDNLKDAVKK	QLQDLYKTRP
	101	EAWREENKKRT	EEAYIEQLGP	KFSILKQKNP	DLINKLVEDS	VLTPHSNTSQ
	151	TSLNNIFNKK	LHVKIENKSH	VAGQVLELTK	MTLKDSLWEP	RRHSDIHMLE
	201	TSDNARIRLN	TKDEKLTVHK	AYQGGADFLF	GYDVRESDPK	ALTFEEKVSG
15	251	QSGVVLERRP	ENLKTLDGRK	LIAAEKADSN	SFAFKQNYRK	GLYELLRLKC
	301	EGGFCLGVQR	LAIPEAAEVL	YAQQAYAANT	LFGLRAADRG	DDVYAADPSR
	351	QKLWLRFIGG	RSHQNIRGGA	AADGRRKGVO	IGGEVFRVQN	EGSRLAIGVM
	401	GGRAGQHASV	NGKGGAAGSY	LHGYYGGVYA	AWHQLRDKQT	GAYLDGWLQY
	451	QRFKHRINDE	NRAERYKTKG	WTASVEGGYN	ALVAEGVVGK	GNNVRFYLPQ
20	501	QAQFTYLVGN	GGFTDSEGTA	VGLLGSGQWQ	SRAGIRAKTR	FALRNGVNLQ
	551	PFAAFNLVLR	SKSFGVEMDG	EKQTLAGRTA	LEGRFGIEAG	WKGHMSARIG
	601	YGKRTDGDKE	AALSLKWLF*			

Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF35 (SEQ ID NO: 452) shows 51.7% identity over a 261aa overlap with a predicted ORF
(ORF35ngh) (SEQ ID NO: 456) from *N. gonorrhoeae*:

	orf35.pep	PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIRG	34
	orf35ngh	FTKVQERDDIAIYAQQAQAANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG	370
30	orf35.pep	GAA-ADGWRKGVQIGGEVFRVQNEGSXLAIGVMGGRAGQHASVNGKG--GAAGSDLYGYG	91
	orf35ngh	KTAPVEGYRKGVLGGGEVFTWQNESNQLSIGLMGGQAEQRFSTFRNPDTDNLTGTGNVKGFG	430
	orf35.pep	GGVYAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAE	151
35	orf35ngh	AGVYATWHQLQDKQTGAYVDSWMQYQRFHRINTEYATERFTSKGITASIEAGYNALLAE	490
	orf35.pep	GIVGKGNNVRFYLPQQAQFTYLGVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRN	211
	orf35ngh	HFTKKGNSLRVYLQPAQLTYLGVNGKFSSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	550
40	orf35.pep	GVNLQPFAAFNVLHRSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA	263
	orf35ngh	GVTFPQPFVAVNSIYQQKPGVEIDGDRRVINNKTVIETQLGVAARIKSHLTLQASFNQRT	610

A partial ORF35ngh nucleotide sequence [SEQ ID 455] (SEQ ID NO: 455) is predicted to encode a protein having partial amino acid sequence [SEQ ID 456] (SEQ ID NO: 456):

45 1 . . KKLDRNRSEY WKEETYHIKS NGRTPNIPA LFPKHPDPF ENINNSKKIS
 51 FYDKEYTEDY LVGFARGFGV EKRNKEEEKP LRQYFKDCVN TENSNDNCK

5
101 ISSFGNYGPI LIKSDIFALA SQIKNSHINS EILSVGNYIE WLRPTLNKLT
151 GWQEHLYAGL DPFHYIEVTD NSHVIGQTID LGALELTNSL WKPRWNSNID
201 YLITKNAEIR FNTKNESLLV KEDYAGGARF RFAYDLKDKV PEIPVLTFEK
251 NITGTSDIIF EGKALDNLKH LDGHQIVKVN DTADKDAFRL SSKYRKGIYT
301 LSLQQRPEGF FTKVQERDDI AIYAQQAQAA NTLFALRLND KNSDIFDRTL
351 PRKGLWLRVI DGHSNQWVQG KTAPVEGYRK GVQLGGEVFT WQNESNQLSI
401 GLMGGQAEQR STFRNPDTDN LTTGNVKGFG AGVYATWHQL QDKQTGAYVD
451 SWMQYQFRFH RINTEYATER FTSKGITASI EAGYNALLAE HFTKKGNLSR
501 VYLQPQAQLT YLGVNGKFSD SENAQVNLLG SRQLQSRVGV QAKAQFAFTN
551 GVTFPFVAV NSIYQQKPGF VEIDGDRRVI NNKTVIETQL GVAARIKSHL
601 TLQASFNRQT SKHHHAKQGA LNLQWTF*

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

15 Example 55

The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 457](#)] ([SEQ ID NO: 457](#)):

20
1 ..GCGGAATATG TTCAGTTCTC TATAGATTG TTCAGTGTGG GTAAATCGGG
51 GGGCGGTATA CCTAAGGCTA AGCCTGTGTT TGATGCGAAA CCGAGATGGG
101 AGGTTGATAG GAAGCTTAAT AAATTGACAA CTCGTGAGCA GGTGGAGAAA
151 AATGTTTCAGG AAACGAGAAG AAGGAGTCAG AGTAGTCAGT TTAAAGCCCA
201 TGCACAACGA GAATGGGAAA ATAAAACAGG GTTAGATTTT AATCATTTTA
251 TAGGTGGTGA TATCAATAAA AAAGGCACAG TAACAGGAGG GCATAGTCTA
301 ACCCGTGGTG ATGTACGGGT GATACAACAA ACCTCGGCAC CTGATAAACA
25 TGGGGT.TTA TCAAGCGACA GTGGAAATTN A

This corresponds to the amino acid sequence [[SEQ ID 458](#); [ORF46](#)] ([SEQ ID NO: 458](#); [ORF46](#)):

30
1 ..AEYVQFSIDL FSVGKSGGGI PKAKPVFDAK PRWEVDRKLN KLTTREQVEK
51 NVQETRRRSQ SSQFKAHAQR EWENKTGLDF NHFIGGDINK KGTVTGGHSL
101 TRGDVRVQQ TSAPDKHGXL SSDSGNX

Further work revealed further partial nucleotide sequence [[SEQ ID 459](#)] ([SEQ ID NO: 459](#)):

35
1 ..GCAGTGTGCC TnCCGATGCA TGCACACGCC TCAnATTGCG CAAACGATTC
51 TTTTATCCGG CAGGTTCTCG ACCGTCAGCA TTTCGAACCC GACGGGAAAT
101 ACCACCTATT CGGCAGCAGG GGGGAACCTG CCGAGCGCCA GTCTCATATC
151 GGATTGGGAA AAATACAAAG CCATCAGTTG GGCAACCTGA TGATTCAACA
201 GGCAGGCCATT AAAGGAAATA TCGGCTACAT TGTCCGCTTT TCCGATCACG
251 GGCACGAAGT CCATTCCCCs TTCGACAACC ATGCCTCACA TTCCGATTCT
40 GATGAAGCCG GTAGTCCCGT TGACGGATTT AGCCTTTACC GCATCCATTG
351 GGACGGATAC GAACACCATC CCGCCGACGG CTATGACGGG CCACAGGGCG
401 GCGGCTATCC CGTCCCAAA GCGCGAGGG ATATATACAG TTACGACATA
451 AAAGGCGTTG CCCAAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC
501 CGGACAACGG CTGCGCGACC GTTTCACAA TGCCGGTAGT ATGCTGACGC
45 AAGGAGTAGG CGACGGATTG AAACGCGCCA CCCGATACAG CCCCAGAGCTG
601 GACAGATCGG GCAATGCCGC CGAAGCCTTC AACGGCACTG CAGATATCGT
651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence [<SEQ ID 460; ORF46-1>] (SEQ ID NO: 460; ORF46-1):

```

5      1  ..AVCLPMHAHA SXLANDSFIR QVLDQRHFEP DGKYHLFGSR GELAERQSHI
      51  GLGKIQSHQL GNLMIQAAAI KGNIGYIVRF SDHGHEVHSP FDNHASHSDS
     101  DEAGSPVDGF SLYRIHWDGY EHHPADGYDG PQGGGYPAKP GARDIYSYDI
     151  KGVAQNIRLN LTDNRSTGQR LADRFRNAGS MLTQGVGDGF KRATRYSPDL
     201  DRSGNAAEAF NGTADIVKNI IGAAGEI

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 (SEQ ID NO: 458) shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) (SEQ ID NO: 462) from *N. gonorrhoeae*:

```

15  orf46.pep          AEYVQFSIDLFSVGKSGGGIPKAKPVFDAKPRWEVDRKLNKLTTTR  45
      |||||||||||||||||||||||||||||||||||||||
  orf46ng          PKTGVFPDGGKGFNFEEKHVKYDTKLDIQELSGGGIPKAKPVFDAKPRWEVDRKLNKLTTTR  217

  orf46.pep          EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIFGGDINKKGTVTGGHSLTRGDV  105
      |||||||||||||||||||||||||||||||||||||||
  orf46ng          EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIFGGDINKKGAVTGGHSLTRGDV  277

20  orf46.pep          RVIQQTSA PDKHGXLSSDSGN  126
      ||||||||||||| |||||
  orf46ng          RVIQQTSA PDKHGVLSSDSGN  298

```

25 A partial ORF46ng nucleotide sequence [<SEQ ID 461>] (SEQ ID NO: 461) is predicted to encode a protein having partial amino acid sequence [<SEQ ID 462>] (SEQ ID NO: 462):

```

30      1  ..RRLKHCCHAR LGS AFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPQC
      51  RTRHRSRQQY LYGSHPHQRD WSCP GKIQLG RHHGTSCRAV ADXRDRICER
     101  EIRRQRQXCR CRLGKIPSL S IPKYPLKLEQ RYKGENITSS TVPPSNGKNV
     151  KLADQRHPKT GVPFDGKGFP NFEKHVKYDT KLDIQELSGG GIPKAKPVFD
     201  AKPRWEVDRK LNKLT TREQV EKNVQETRRR SQSSQFKAHA QREWENKTGL
     251  DFNHFIFGGDI NKKGA VTGGH SLTRGDV RVI QQTSA PDKHG VLSSDSGN*

```

Further work revealed the complete gonococcal DNA sequence [<SEQ ID 463>] (SEQ ID NO: 463):

```

35      1  TTGGGCATTT CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG
      51  CCTGCCGATG CATGCACACG CCTCAGATTT GGcaAACGAT CCCTTTATCC
     101  GgCaggttcT CGaccGTCAG CATTTCGaac ccgacggGAa ATACCaCCTA
     151  TtcggCaGCA GGGGGGAGCT TgccnagcGC aacggccATa tcggattggG
     201  aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcggccg
40    251  ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa
     301  ttccattcgc ctttcGAcAa ccaTGCCTCA CATCCGATT CTGACGAAGC

```

-355-

5
10
15
20
25
30

```

351  CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT
401  ACGAACACCA TCCCGCCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT
451  CCGGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAAGGCGT
501  TGCCCAAAAT ATCCGCCTCA ACCTGACCGA CAACCGCAGC ACCGGACAAC
551  GGCTTGCCGA CCGTTTCCAC AATGCCGGCG CTATGCTGAC GCAAGGAGTA
601  GGCACGGGAT TCAAACGCGC CACCCGATAC AGCCCCGAGC TGGACAGATC
651  GGGCAATGCG gccGAAGCCT TCAACGGCAC TGCAGATATC GTCAAAAACA
701  TCATCGGCGC GGCAGGAGAA ATTGTCGGCG CAGGCGATGC CGTGCagGGT
751  ATAAGCGAAG GCTCAAACAT TGCTGTGATG CACGGCTTGG GTCTGCTTTC
801  CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC
851  TCAAAGACTA TGCCGCAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC
901  AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA TGGCAGCCAT
951  CCCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA
1001 TCACGGCACA TCCTGTCAAG CGGTCGAGA TGGGCGCGAT CGCATTGCCG
1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA
1101 ATACCCGTCC CTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC
1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC
1201 AAAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT
1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA
1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAAAAATAA
1501 ACAGGGTTAG ATTTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG
1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC
1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGAA
1651 ATTAAAAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA
1701 AGTGATGACC AAGCACACCA TGTTCCCAAA AGATTGGGAT GAGGCTAGAA
1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
1801 AATAAATGGC AGGGTACAAG TAAATCGGGT ATTAAATAG AAGGATTTAC
1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

```

This corresponds to the amino acid sequence [<SEQ ID 464; ORF46ng-1>] (SEQ ID NO: 464; ORF46ng-1):

35
40
45

```

1  LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKYHL
51  FGSRGELAXR NGHIGLNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSDHGKH
101 FHSPPDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD GYDGPQGGGY
151 PAPKGARDIY SYDIKGVAQN IRLNLTNRS TGQRLADRFH NAGAMLTQGV
201 GDGFKRATRY SPELDRSGNA AEAFTGTADI VKNIIGAAGE IVGAGDAVQG
251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
301 NAAQIEAVS NIFMAAIPK GIGAVRGKYG LGGITAHVPK RSQMGAIALP
351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
401 KNVKLADQRH PKTGVFPDGK GFPNFEKHVK YDTKLDIQEL SGGGIPKAKP
451 VFDAKPRWEV DRKLNKLTTT EQVEKNVQET RRRSQSSQFK AHAQREWENK
501 TGLDFNHFIC GDINKKGTVT GGHSLTRGDV RVIQQTAPD KHGVYQATVE
551 IKKPDGSWEV KTKKGGKVM T KHTMFPKDW EEARIRAEVTS AWESRIMLKD
601 NKWQGTSKSG IKIEGFTEPN RTAYPIYE*

```

ORF46ng-1 (SEQ ID NO: 464) and ORF46-1 (SEQ ID NO: 460) show 94.7% identity in 227 aa overlap:

55

```

          10      20      30      40
orf46-1.pep      AVCLPMHAHASXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER
                  |||||
orf46ng-1      LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR
          10      20      30      40      50      60

```

-356-

		50	60	70	80	90	100
	orf46-1.pep	QSHIGLGKIQSHQLGNLMIQQA	AIKGNIGYIVRFS	DHGHVHSPFDN	HASHSDSDEAGSP		
5	orf46ng-1	NGHIGLGNIQSHQLGHLMIQQA	AVEGNIGYIVRFS	DHGHKHFHSPFDN	HASHSDSDEAGSP		
		70	80	90	100	110	120
		110	120	130	140	150	160
	orf46-1.pep	VDGFSLYRIHWDGYEHHPADGYDGPQGGGY	PAPKGARDIYSYDI	KGVAQNIRLNLT	DNRS		
10	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGGY	PAPKGARDIYSYDI	KGVAQNIRLNLT	DNRS		
		130	140	150	160	170	180
		170	180	190	200	210	220
	orf46-1.pep	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSP	ELDRSGNAAEFNGTADIVKNIIGAAGE				
15	orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSP	ELDRSGNAAEFNGTADIVKNIIGAAGE				
		190	200	210	220	230	240
	orf46-1.pep	I					
20	orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENK	MARINDLADMAQLKDYAAAAIRDWAVQNP				
		250	260	270	280	290	300

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 (SEQ ID NO: 464) shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) (SEQ ID NO: 466) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
25	orf46a.pep	LGISRKISLILSILAVCLPMHAHASD	LANDSFIRQVLDRQHFEPD	GKYHLFGSRGELAER			
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASD	LANDPFIRQVLDRQHFEPD	GKYHLFGSRGELAXR			
		10	20	30	40	50	60
		70	80	90	100	110	120
30	orf46a.pep	SGHIGLGNIQSHQLGNLFIQQA	AIKGNIGYIVRFS	DHGHVHSPFDN	HASHSDSDEAGSP		
	orf46ng-1	NGHIGLGNIQSHQLGHLMIQQA	AVEGNIGYIVRFS	DHGHKHFHSPFDN	HASHSDSDEAGSP		
		70	80	90	100	110	120
		130	140	150	160	170	180
35	orf46a.pep	VDGFSLYRIHWDGYEHHPADGYDGPQGGGY	PAPKGARDIYSYDI	KGVAQNIRLNLT	DNRS		
	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGGY	PAPKGARDIYSYDI	KGVAQNIRLNLT	DNRS		
		130	140	150	160	170	180
		190	200	210	220	230	240
40	orf46a.pep	TGQRLVDRFHNTGSMLTQGVGDGFKRATRYSP	ELDRSGNAAEFNGTADIVKNIIGAAGE				
	orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSP	ELDRSGNAAEFNGTADIVKNIIGAAGE				
		190	200	210	220	230	240
		250	260	270	280	290	300
45	orf46a.pep	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENK	MARINDLADMAQLKDYAAAAIRDWAVQNP				

-357-

	orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP	250	260	270	280	290	300
5	orf46a.pep	NAAQGIEAVSNIFTAVIPVKGIGAVRGKYGLGGITAHVPKRSQMGEIALPKGKSAVSDNF	310	320	330	340	350	360
	orf46ng-1	NAAQGIEAVSNIFMAAIIPIKGIGAVRGKYGLGGITAHVPKRSQMGAIALPKGKSAVSDNF	310	320	330	340	350	360
10	orf46a.pep	ADAAYAKYPSPYHSRNIIRSNIQRYGKENITSSTVPPSNGKNVKLANRHPKTKVPFDGK	370	380	390	400	410	420
	orf46ng-1	ADAAYAKYPSPYHSRNIIRSNIQRYGKENITSSTVPPSNGKNVKLADQRHPKTKVPFDGK	370	380	390	400	410	420
15	orf46a.pep	GFPNFEKDVKYDTRINTAVPQVN---PIDEPVFN--PKGSVGSASHSWSITARIQYAKLP	430	440	450	460	470	
	orf46ng-1	GFPNFEKHVKYDTKLD--IQELSGGGIPKAKPVFDAKPRWEVDRKLN-KLTREQVEKQNV	430	440	450	460	470	
20	orf46a.pep	RQGRIRYIPPKNYSPSAPLPKGPNNGYLDKFGNEWTKGPSRTKGQEFEDWDVQLSKTGREQ	480	490	500	510	520	530
	orf46ng-1	QETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDVRVVIQQTS	480	490	500	510	520	530

The complete length ORF46a DNA sequence [<SEQ ID 465>] (SEQ ID NO: 465) is:

25	1	TTGGGCATTT	CCCGCAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATTT	GGCAAACGAT	TCTTTTATCC
	101	GGCAGGTTCT	CGACCGTCAG	CATTTCGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGGGGAAC	TGCCGAGCGC	AGCGGTCATA	TCGGATTGGG
30	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCCAG	CAGGCGGCCA
	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGGCACGAA
	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
35	451	CCCGCTCCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAAGGCGT
	501	TGCCCCAAAT	ATCCGCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
	551	GGCTTGTGCA	CCGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
40	701	TCATCGGCGC	GGCAGGAGAA	ATTGTCCGGC	CAGGCGATGC	CGTGCAGGGT
	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTTATG	CACGGCTTGG	GTCTGCTTTC
	801	CACCGAAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC
	851	TCAAAGACTA	TGCCGAGCA	GCCATCCGCG	ATTGGGCAGT	CCAAAAACCC
	901	AATGCCGCAC	AAGGCATAGA	AGCCGTCAGC	AATATCTTTA	CGGCAGTCAT
45	951	CCCCGTCAA	GGGATTGGAG	CTGTTCTGGG	AAAATACGGC	TTGGGCGGCA
	1001	TCACGGCACA	TCCTGTCAAG	CGGTCGCAGA	TGGGCGAGAT	CGCATTGCCG
	1051	AAAGGGAAAT	CCGCCGTCAG	CGACAATTTT	GCCGATGCGG	CATACGCCAA
	1101	ATACCCGTCC	CCTTACCATT	CCCGAAATAT	CCGTTCAAAC	TTGGAGCAGC
	1151	GTTACGGCAA	AGAAAACATC	ACCTCCTCAA	CCGTGCCGCC	GTCAAACCGA
	1201	AAGAATGTGA	AACTGCAAAA	CAAAACGCCAC	CCGAAGACCA	AAGTGCCGTT
50	1251	TGACGGTAAA	GGGTTTCCGA	ATTTTGAAAA	AGACGTAAAA	TACGATACGA
	1301	GAATTAATAC	CGCTGTACCA	CAAGTGAATC	CTATAGATGA	ACCCGTCTTT
	1351	AATCCTAAAG	GTTCTGTCGG	ATCGGCTCAT	TCTTGGTCTA	TAACTGCCAG
	1401	AATTCATAC	GCAAAATTAC	CAAGGCAAGG	TAGAATCAGA	TATATCCAC
	1451	CTAAAAATTA	CTCTCCTTCA	GCACCGCTAC	CAAAAGGACC	TAATAATGGA

1501 TATTTGGATA AATTTGGTAA TGAATGGACT AAAGGTCCAT CAAGAACTAA
 1551 AGGTCAAGAA TTTGAATGGG ATGTTCAATT GTCTAAAACA GGAAGAGAGC
 1601 AACTTGGATG GGCTAGTAGG GATGGTAAGC ATTTAAATAT ATCAATTGAT
 1651 GGAAAGATTA CACACAAATG A

5

This corresponds to the amino acid sequence [<SEQ ID 466>] (SEQ ID NO: 466):

10 1 LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRO HFEPDGKYHL
 51 FGSRGELAER SGHIGLGNIG SHQLGNLFIQ QAAIKGNIGY IVRFSDHGHE
 101 VHSFPDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD GYDGPQGGGY
 151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLVDRFH NTGSMLTQGV
 201 GDGFKRATRY SPELDRSGNA AEAFNGTADI VKNIIGAAGE IVGAGDAVQG
 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQGIEAVS NIFTAVIPVK GIGAVRGKYG LGGITAHVPK RSQMGEIALP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
 15 401 KNVKLANKRH PKTKVPFDGK GFPNFEKDVK YDTRINTAVP QVNPIDEPVF
 451 NPKGSVGS AH SWSITARIQY AKLPRQGRIR YIPPKNYS PS APLPKGPNNG
 501 YLDKFGNEWT KGPSRTKGQE FEWDVQLSKT GREQLGWASR DGKHLNISID
 551 GKITHK*

20 Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 56

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 467>] (SEQ ID NO: 467):

25 1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTC
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT
 201 CAAAATTGCC GGCCTATTGG CGTTTGGCT GCGGTTTTTG TTTGACGGGC
 30 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCTTATC AGATAATGAC
 351 CGGGCTG...

This corresponds to the amino acid sequence [<SEQ ID 468; ORF48>] (SEQ ID NO: 468; ORF48):

35 1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWVLAL LTATARPIVN
 51 LDYLPALLI ALPWRVFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
 101 NLVPFILTAP APYQIMTGL...

Further work revealed the complete nucleotide sequence [<SEQ ID 469>] (SEQ ID NO: 469):

40 1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTC
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT

5
10
15
20
25

```

151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT
201 CAAAATTGCC GGCGTATTGG CGTTTGGCT GGCGGTTTGG TTGACGGGC
251 TGATGATGGT GATCCAACTC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
401 CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTTAC CGGCCATTG AGTTACTACG ACCGGGGTGC
501 GATGGCCAAT ATCTTCGGCG CAAACAACCT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGCCAATCCC GAACTTCAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCCG
901 TGCCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCATGCA
951 CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACTGC
1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTGCGCG AAGTGTCGGC
1101 ATTTTTCAAA AAACACGACA AGGACTGTT TTA CTGGATG ACCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC CTCTGCCGCA ATTCAGCCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
1351 AACCTCAATG AAACCTTCCG CTACCTCAA CAGGGGCACG TCGCCTGGCT
1401 GAACTTCAA ATCAAATA

```

This corresponds to the amino acid sequence [<SEQ ID 470; ORF48-1>] (SEQ ID NO: 470; ORF48-1):

30
35
40

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWVLAL LTATARPIVN
51 LDYLPALLI ALPWRPVKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPPILTAP APYQIMTGLL LLYMLAMPFV LQKAAAKTDF RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANNFYFA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNAATFAKL
251 LAQKDRFSVW ESGSFPIGA TVEGEMRELC AYGGLRGFAL RRAPDEKFAF
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKKTG
351 AIFGGVCDSE LFGEVSAPFK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPPVG
451 NLNETFRYLK QGHVAWLNFK IK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF48 (SEQ ID NO: 468) shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) (SEQ ID NO: 472) from strain A of *N. meningitidis*:

45

```

          10      20      30      40      50      60
orf48.pep MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNL DYLPALLI
          |||||
orf48a    MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNAVFWVLALLTATARPIVNLXYLPAALLI
          10      20      30      40      50      60

```

-360-

		70	80	90	100	110	119
orf48.pep		ALPWR	FKIAGV	LAFWL	AVLFDGL	MMVIQL	FPFMDLIGAINLV
orf48a		ALPWR	XVKIXG	VLAXWL	AVLFDGL	MMVIQL	FPFMDLIGAINLV
		70	80	90	100	110	120
orf48a		LLYML	AMPFVL	QKAAAK	TDFRHI	AACA	AVVVAAGYFTGHL
		130	140	150	160	170	180

The complete length ORF48a nucleotide sequence [<SEQ ID 471>] (SEQ ID NO: 471) is:

```

10      1  ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCCT
      51  GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCCAATGCGG
      101  TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
      151  TTGGANTACC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTNTCGT
      201  CAAAATTGNC GCGGTATTGG CGTNTTGGCT GGCGGTTTGT TTTGACGGGC
      15      251  TGATGATGGT GATCCAACCT TTCCCTTTTA TGGATCTCAT CGGCGCCATC
      301  AACCTCGTCC CCTTCATCNT GACCGCCCCC GCCCTTTATC AGATAATGAC
      351  CGGGCTGTTA CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
      401  CCGCCGCCAA AACCGACTTC CGACACATTG CCGCCTGTGC CGCCGTTGTG
      451  GTGGCAGCCG GCTATTTTAC CGGCCATTGT AGTTANTACG ACCGGGGGCG
      20      501  GATGCCCAAT ATCTTCGGCG CAAACAACCT CTATTACGCC AAAAGTCAGG
      551  CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
      601  GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
      651  TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
      701  GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
      25      751  CTGGCGCAAA AAGANCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
      801  CATCGGCGCG ACGATCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
      851  GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC
      901  TGCCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
      951  CGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
      30      1001  GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
      1051  GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTGCGCG AAGTGTGCGC
      1101  ANTTTTCAAA AAACACGACA AGGACTGTT TTA CTGGATG ACGCTGACCA
      1151  GCCACGCCGA CTATCCCGAA TCNGACATT TCAACACAG GCTCAAATGC
      1201  ACCGAATATG GCCTGCCCGC CGAAACCGAC NTCTGCCGCA ATTTACAGCT
      35      1251  GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
      1301  TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
      1351  AACCTCAATG AAACCTTCCG CTACCTCAAA CAGGGGCACG TCGNCTGGCT
      1401  GAACTTCAAA ATCAAATAA

```

40 This encodes a protein having amino acid sequence [<SEQ ID 472>] (SEQ ID NO: 472):

```

      1  MNIHTLLSKQ WTLPPFLPKR LLLSLILLX PNAVFVWLAL LTATARPIVN
      51  LXYLPAALLI ALPWRXVKIX GVLAXWLAVL FDGLMMVIQL PPFMDLIGAI
      101  NLVPFIXTAP ALYQIMTGLL LLYMLAMPFV LQKAAAKTDF RHIAACAAYV
      151  VAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
      45      201  VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNATFAKL
      251  LAQKKRFSVW ESGSFFFIGA TIEGEMRELC AYGGRLGFAL RRAPDEK FAR
      301  CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGK KTC
      351  AIFGGVCDSE LFGEVSAXFK KHDKGLFYWM TLTSHADYPE SDIFNHLKC
      401  TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVVG
      50      451  NLNETFRYLK QGHVXWLNFK IK*

```

ORF48a (SEQ ID NO: 472) and ORF48-1 (SEQ ID NO: 470) show 96.8% identity in 472 aa overlap:

-361-

		10	20	30	40	50	60
	orf48a.pep	MNIHTLLSKQWTLPPFLPKRLLSLLILLXPNVFWVLALLTATARP	IVNLYLPAALLI				
5	orf48-1	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARP	IVNLDYLPALLI				
		10	20	30	40	50	60
	orf48a.pep	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLV	VPFIXTAPALYQIMTGLL				
10	orf48-1	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLV	VPFIXTAPALYQIMTGLL				
		70	80	90	100	110	120
	orf48a.pep	LLYMLAMPFVLQKAAAKTDFRHIACA	AVVVAAGYFTGHLSXYDRGRMANIFGANNFY				
15	orf48-1	LLYMLAMPFVLQKAAAKTDFRHIACA	AVVVAAGYFTGHLSXYDRGRMANIFGANNFY				
		130	140	150	160	170	180
	orf48a.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQ	QRAATHLNEPKSQKILFIVAESWGLPANP				
20	orf48-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQ	QRAATHLNEPKSQKILFIVAESWGLPANP				
		190	200	210	220	230	240
	orf48a.pep	ELQNATFAKLLAQKXRFVSWESGSFF	FIGATIEGEMRELCAYGGLRGFALRRAPDEK				
25	orf48-1	ELQNATFAKLLAQKXRFVSWESGSFF	FIGATIEGEMRELCAYGGLRGFALRRAPDEK				
		250	260	270	280	290	300
	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWY	PRAGFQEI	KTAENLIGKKTCAIFGGVCDSE			
30	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWY	PRAGFQEI	KTAENLIGKKTCAIFGGVCDSE			
		310	320	330	340	350	360
	orf48a.pep	LFGEVSAXFKKHDKGLFYWMTLTSHADY	PESDIFNHRLKCTEYGLPAETDXCRNFS				
35	orf48-1	LFGEVSAXFKKHDKGLFYWMTLTSHADY	PESDIFNHRLKCTEYGLPAETDXCRNFS				
		370	380	390	400	410	420
	orf48a.pep	FFDQLADLIQRPEMKGTEVIVGDHPP	PVGNLNETFRYLKQGHVXWLNFKIKX				
40	orf48-1	FFDQLADLIQRPEMKGTEVIVGDHPP	PVGNLNETFRYLKQGHVXWLNFKIKX				
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF48. (SEQ ID NO: 468) shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) (SEQ ID NO: 474) from *N. gonorrhoeae*:

45	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARP	IVNLDYLPALLI	60
	orf48ng	MNIHALLSEQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARP	IVNLDYLPALLI	60

orf48.pep	ALPWRFVKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	119
orf48ng	ALPWRFVKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL	120

- 5 The ORF48ng nucleotide sequence [<SEQ ID 473>] (SEQ ID NO: 473) was predicted to encode a protein having amino acid sequence [<SEQ ID 474>] (SEQ ID NO: 474):

1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	PNAVFWVLAL	LTATARPIVN
51	LDYLPAAALLI	ALPWRFVKIA	GVLAFWPAVL	FDGLMMVIQL	FPFMDLIGAI
101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LQKAAVKTDF	RHIAVCAAVV
151	AAARYFTGPF	ELLRTGGRWQ	YVQHRRLLS	GSRASFRRRQ	KADVLRLRGN
201	PYASMGNGG.				

Further work identified the complete gonococcal DNA sequence [<SEQ ID 475>] (SEQ ID NO: 475):

1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCTCT
51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCGG
101	TGTTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCCCGCC	GATTGTCAAT
151	TTGGACTACC	TTCCCGCCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
201	CAAAATTGCC	GGCGTATTGG	CGTTTTGGCC	GGCGGTTTTG	TTTGACGGGC
251	TGATGATGGT	GATCCAACCT	TTCCCTTTTA	TGGACCTCAT	CGGCGCCATC
301	AACCTCGTCC	CCTTCATCCT	GACCGCCCCC	GCCCCTTATC	AGATAATGAC
351	CGGGCTGTTG	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAAAAAG
401	CCGCCGTCOA	AACCGACTTC	CGACACATTG	CCGTCTGTGC	CGCCGTTGTG
451	GCGGCAGCCG	GCTATTTTAC	CGGCCATTTG	AGTTACTACG	ACCGGGGGCG
501	GATGGCCAAT	ATCTTCGGCG	CAAACAACCT	CTATTACGCC	aAAAGTCAGG
551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGgcctg
601	GTCGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CCGCCACGCG
651	GCTGAGTGAG	CCGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCACTTT	TGCCAAACTG
751	CTGGCGCAAA	AAGACCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCTTT
801	CATCGGCGCG	ACGGTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
851	GTTTGC GCGG	GTTTCGACTG	CGCCGCGCGC	CCGACGAAAA	ATTGCCCCGC
901	TGCCTCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
951	CGGCGCGGGT	AGTTCGCTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
1001	GCTTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTCGGCG	AAGTGTGCGG
1101	ATTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
1151	GCCACGCCGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
1201	ACCGAATACG	GCCTGCCCCG	CGAAACCGAC	CTCTGCCGCA	ATTTCAGCCT
1251	GCACACCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCCGAAA
1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCCCGTCGGC
1351	AACCTCAATG	AAACCTTCCG	CTACCTCAAA	CAGGGACACG	TCGCCTGGCT
1401	GCACTTCAAA	ATCAAATAA			

- 45 This encodes a protein having amino acid sequence [<SEQ ID 476; ORF48ng-1>] (SEQ ID NO: 476; ORF48ng-1):

1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	PNAVFWVLAL	LTATARPIVN
51	LDYLPAAALLI	ALPWRFVKIA	GVLAFWPAVL	FDGLMMVIQL	FPFMDLIGAI
101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LQKAAVKTDF	RHIAVCAAVV
151	AAAGYFTGHL	SYYDRGRMAN	IFGANNFYYA	KSQAMLYTVS	QNADFITAGL
201	VDPVFLPLGN	QQRATRLSE	PKSQKILFIV	AESWGLPGNP	ELQNATFAKL

5

251 LAQKDRFSVW ESGSFPPFIGA TVEGEMRELC AYGGLRGFAL RRAPDEK FAR
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQKIKT AENLIGK KTC
351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHR LKC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPP PVG
451 NLNETFRYLK QGHVAWLHFK IK*

ORG48ng-1 (SEQ ID NO: 476) and ORF48-1 (SEQ ID NO: 470) show 97.9% identity in 472 aa overlap:

10	orf48-1.pep	10 20 30 40 50 60	MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARPIVNLDYLPALLI
	orf48ng-1	10 20 30 40 50 60	MNIHALLSEQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARPIVNLDYLPALLI
15	orf48-1.pep	70 80 90 100 110 120	ALPWRFVKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL
	orf48ng-1	70 80 90 100 110 120	ALPWRFVKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL
20	orf48-1.pep	130 140 150 160 170 180	LLYMLAMPFVLQKAAAKTDFRHIIVCAAVVAAAGYFTGHLSYYDRGRMANIFGANNFYA
	orf48ng-1	130 140 150 160 170 180	LLYMLAMPFVLQKAAVKTDPRHIIVCAAVVAAAGYFTGHLSYYDRGRMANIFGANNFYA
25	orf48-1.pep	190 200 210 220 230 240	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLNEPKSQKILFIVAESWGLPANP
	orf48ng-1	190 200 210 220 230 240	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATRLSEPKSQKILFIVAESWGLPGNP
30	orf48-1.pep	250 260 270 280 290 300	ELQNATFAKLLAQKDRFSVWESGSFPPFIGATVEGEMRELCAYGGLRGFALRRAPDEK FAR
	orf48ng-1	250 260 270 280 290 300	ELQNATFAKLLAQKDRFSVWESGSFPPFIGATVEGEMRELCAYGGLRGFALRRAPDEK FAR
35	orf48-1.pep	310 320 330 340 350 360	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEI KTAENLIGKKTCAIFGGVCDSE
	orf48ng-1	310 320 330 340 350 360	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQKIKTAENLIGKKTCAIFGGVCDSE
40	orf48-1.pep	370 380 390 400 410 420	LFGEVSAFFKKHDKGLFYWM TLTSHADYPESDIFNHR LKCTEYGLPAETDLCRNFSLHTQ
	orf48ng-1	370 380 390 400 410 420	LFGEVSAFFKKHDKGLFYWM TLTSHADYPESDIFNHR LKCTEYGLPAETDLCRNFSLHTQ
45	orf48-1.pep	430 440 450 460 470	FFDQLADLIQRPEMKGTEV IIVGDHPPPVGNLNETFRYLKQGHVAWLNFKIKX
	orf48ng-1	430 440 450 460 470	FFDQLADLIRPEMKGTEV IIVGDHPPPVGNLNETFRYLKQGHVAWLHFKIKX

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 57

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 477] (SEQ ID NO: 477):

```

1  ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCACAAA TCATCATCGT
51  TACTTTGAGT ATCGCCACGC TTGCCGCCGC CGGCATCGCT ATGTCGCGCG
10  101  GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
151  GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
201  AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
251  CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
301  AGTGGCGGTTT TGGCTTTGGT TTTCCTTGCA CTGGGCGC.G TAGCGCCGAA
15  351  CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAATAT AACGGGCAAT
401  TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence [SEQ ID 478; ORF53] (SEQ ID NO: 478; ORF53):

```

20  1  ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMQSD FIEPTPWTLA
51  GLGFLIALMG WMPAPIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
101  SAVLALVFLA LGXVAPNGNG XTVQMAGGKY NGQLINMYA..

```

Further work revealed the complete nucleotide sequence [SEQ ID 479] (SEQ ID NO: 479):

```

25  1  ATGTCCGAAC AACATATTTT GACTTGGAAA AGTAAATCA ACGCATTGGG
51  TCCGGGGATC ATGATGGCTT CGGCGGCGGT CGGCGGTTGC CACCTGATTG
101  CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
151  ATCCTGACCA ACCTCTTCAA ATACCGGTTT TTCCGTTTCA GCGCGCATTG
201  CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
30  251  GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
301  AACCGGGGCG CGGTGCGCAT TGTAAACGCC GCCATCGTCA AAATGGCGAT
351  TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
401  CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
451  TCCAAATATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCCGG
35  501  CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
551  CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601  ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
651  AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTGATTTTCA
701  ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTTT CCTTGCACTG
40  751  GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
801  CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT
851  GGTTCGCGCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCACG
901  ACGATTACCG TCGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG
951  CCTGCTGCGC GGAAGAGACA AAACGGGCAA CGCCGAATTC TTTGCCTGGA
45  1001 ATATTTGGGT GCGGGCAGC GGTTCGCGG TGATTTTCTG GTTTGACGGC

```

1051	GTAATGGCGA	ATCTGCTCAA	ATTTCGCGATG	ATTGCCGCTT	TTGTGTCCGC
1101	CCCTGTGTTT	GCTTGGCTGA	ATTACCGTTT	GGTTAAAGGT	GATGAAAAAC
1151	ACAAACTCAC	ATCAGGTATG	AATGCCCTTG	CATTGGCAGG	CTTGATTTAT
1201	CTGACCGGTT	TTACCGTTTT	GTTCTTATTG	AATTTGCCGG	GAATGTTCAA
1251	ATGA				

This corresponds to the amino acid sequence [SEQ ID 480; ORF53-1] (SEQ ID NO: 480; ORF53-1):

1	MSEQHISTWK	SKINALGPGI	MMASAAVGG	HLIASTQAGA	<u>LYGWQIALII</u>
51	<u>ILTNLFKYPF</u>	FRFSAHYTLD	TGKSLIEGYA	EKSRVYLWVF	<u>LILCILSATI</u>
101	NAGAVAIVTA	AIVKMAIPSL	MFDAGTVAAL	IMASCLIIIV	SGRYRALDRV
151	SKIIIVTLSI	<u>ATLAAAGIAM</u>	SRGMQMOSDF	IEPTPWTLAG	LGFLIALMGW
201	MPAPIEISAI	NSLWVTEKQR	INPSEYRDI	FDNFVGYIAS	<u>AVLALVFLAL</u>
251	<u>GAFVQYGN</u>	AVQMAGGKYI	QLINMYAVT	IGGWSRPLVA	<u>FIAFACMYGT</u>
301	<u>TITVDGYAR</u>	IAIEPVRLLR	GKDKTGNAEF	FAWNLWVAGS	<u>GLAVIFVFDG</u>
351	<u>VMANLLKFAM</u>	<u>IAAFVSAPVF</u>	AWLNRYLVKG	DEKHKLTSGM	<u>NALALAGLIY</u>
401	LTGFTVLFL	LLNLAGMFK*			

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF53 (SEQ ID NO: 478) shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) (SEQ ID NO: 482) from strain A of *N. meningitidis*:

25 orf53.pep VSGRYRALDRVSKIIIVTLSIATLAAAGIA 10 20 30
|||

25 orf53a AAIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIA 110 120 130 140 150 160

30 orf53.pep MSRGMQMQSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG 40 50 60 70 80 90
|||

30 orf53a MSRGMQMQSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG 170 180 190 200 210 220

35 orf53.pep IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQMAGGKYNGQLINMYA 100 110 120 130 139
|||: |||: |||

35 orf53a IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLV 230 240 250 260 270 280

orf53a AFIAFACMYGTTITVVDGYARAIAPVRLLRGDKTGNAEFFAWNIWVAGSGLAVIFWFD 290 300 310 320 330 340

40 The complete length ORF53a nucleotide sequence [<SEQ ID 481>] (SEQ ID NO: 481) is:

1 ATGTCCGAAC AACATATTTC GACTTGAAAA AGTAAAATCA ACGCATTGGG
51 ACCGGGGATT ATGATGGCTT CGGCGGCGGT CGGCGGTTCG CACTTGATTG
101 CCTCGACGCA GGC GGCGCG CTTTACGGCT GGAGATCGC GCTCATCATC
151 ATCCTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGCATTA

-366-

5
10
15
20

```

201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
251 GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
301 AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA AAATGGCGAT
351 TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
401 CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
451 TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCCGG
501 CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601 ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
651 AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTCa
701 ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGTTTTT CCTTGCACTG
751 GGC GCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT
851 GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCACG
901 ACGATTACCG TTGTGGACGG CTATGCCCGT GCCATTGCCG AACCCTGCG
951 CTGCTGCGC GGAAGAGACA AAACGGGCAA CGCCGAATTC TTTGCTGGA
1001 ATATTGGGT GCGGGGCGC GGTGGGCGG TGATTTCTG GTTTGACGGC
1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
1101 CCCTGTGTTT GCCTGGCTGA ATTACGTTT GGTCAAAGGT GATGAAAAAC
1151 ACAAACACAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTAT
1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
1251 ATGA

```

This encodes a protein having amino acid sequence [SEQ ID 482] (SEQ ID NO: 482):

25
30

```

1  MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASTQAGA LYGWQIALII
51  ILTNLFKYPF FRFSAHYTLD TGKSLIEGYA EKSrvYLWVF LILCILSATI
101 NAGAVAIvTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
151 SKIIIVTISI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW
201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL
251 GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIWVAGS GLAVIFWFDG
351 VMANLLKFAM IAAFVSAPVF AWLNYRLVKG DEKHKLTSGM NALALAGLIY
401 LTGFTVLFLN LLAGMFK*

```

35 ORF 53a (SEQ ID NO: 482) shows 100.0% identity in 417 aa overlap with ORF53-1 (SEQ ID NO: 480):

40
45
50

```

          10      20      30      40      50      60
orf53a.pep MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYPF
          |||||
orf53-1     MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYPF
          10      20      30      40      50      60

          70      80      90      100     110     120
orf53a.pep FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIvTAAIVKMAIPSL
          |||||
orf53-1     FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIvTAAIVKMAIPSL
          70      80      90      100     110     120

          130     140     150     160     170     180
orf53a.pep MFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTISIATLAAAGIAMSRGMQMOSDF
          |||||
orf53-1     MFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTISIATLAAAGIAMSRGMQMOSDF
          130     140     150     160     170     180

          190     200     210     220     230     240

```

-367-

	orf53a.pep	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEQRINPSEYRDGIFDFNVGYIAS
	orf53-1	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEQRINPSEYRDGIFDFNVGYIAS
		190 200 210 220 230 240
5	orf53a.pep	250 260 270 280 290 300 AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
	orf53-1	AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
		250 260 270 280 290 300
10	orf53a.pep	310 320 330 340 350 360 TITVVDGYARAI AEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM
	orf53-1	TITVVDGYARAI AEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM
		310 320 330 340 350 360
15	orf53a.pep	370 380 390 400 410 IAAFVSAPVFAWLNRYLVKGD EKHKL TSGMNALALAGLIYLTGFTVFLNLAGMFKX
	orf53-1	IAAFVSAPVFAWLNRYLVKGD EKHKL TSGMNALALAGLIYLTGFTVFLNLAGMFKX
		370 380 390 400 410

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 (SEQ ID NO: 478) shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) (SEQ ID NO: 484) from *N. gonorrhoeae*:

25	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIA	91
	orf53.pep	MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEQRINPSEYRDG	90
	orf53ng	MSRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEQRINPSEYRDG	151
30	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQMAGGKYNGQLINMYA	139
	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMGGGKYIGQLINMYAVTIGGGSRPLV	211

An ORF53ng nucleotide sequence [<SEQ ID 483>] (SEQ ID NO: 483) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 484):

35	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMOP
	101	DFIEPTPWT LAGLGFLIALM GWMPAPIEIS AINSLWVTEK QRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
	201	VTIGGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRGKDKTARP
40	251	IVLLEKLGGR HRFRGDFLV*

Further analysis revealed further partial DNA gonococcal sequence [<SEQ ID 485>] (SEQ ID NO: 485):

1 ..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
 51 CGCCACGATT AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA
 101 AAATGGCGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG
 151 ATTATGGCAT CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT
 5 201 GGATCGTGTT TCCAAAATCA TCATTGTTAC TTTGAGCATC GCCACGCTTG
 251 CCGCCGCCGG CATCGCTATG TCGCGCGGTA TGCAGATGCA GCCCGATTTT
 301 ATCGAGCCGA CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT
 351 GATGGGCTGG ATGCCCCGCG CGATCGAAAT TTCCGCCATC AATTCTTTGT
 401 GGGTAACCGA AAAACAACGC ATCAATCCTT CTGAATACCG CGACGGGATT
 10 451 TTCGATTCA ACCTCGGTTA TATCGCagT GCGGTTTTGG CTTTGTTTTT
 501 CCTTGCACTG GCGCGTTTTG TGCAATACGG CAACGCGCAA GCAGTGCAGA
 551 TGGCGGGCGG CAAATATATC GGGCAATTGA TTAATATGTA TGCCGTAACC
 601 ATCGGCGGCT GGTCTCGTCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT
 651 GTACGGCACG ACGATTACCG TTGTGGACGG TTATGCGCGT GCCATTGCCG
 15 701 AACCCGTGCG CCTGCTGCGC GGCAGGGATA AAACCGGCAA CGCCGAGTTG
 751 TtTgCcTGGa ATATTTGGGT GCGGGCAGC GGTtTGGCGG TGATTTTCTG
 801 GTTTGACggc gcaaTGGCgG AActgcTCAa ATTTGCGATG ATtggcgCCT
 851 TTGTGTCCGC CCCTGTGTTT GCCTGGCTCA ACTACCGCCT CGTCAAAGGG
 901 GACAAACGCC ACAGGCTTAC CGCCGGTATG AACGCCCTTG CCATTGTCGG
 20 951 CCTGCTCTAC CTGGCCGGGT TTGCCGTTTT GTTCCTGTTG AACCTTACCG
 1001 GACTTTTGGC ATAG

This corresponds to the amino acid sequence [[SEQ ID 486](#); ORF53ng-1] ([SEQ ID NO: 486](#); [ORF53ng-1](#)):

25 1 ..KKSCVYLWVF LILCIASATI NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL
 51 IMASCLIIILV SGRYRALDRV SKIIIVTLSI ATLAAAGIAM SRGMQMOPDF
 101 IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEKQR INPSEYRDGI
 151 FDFNVGYIAS AVLALVFLAL GAFVQYNGE AVQMAGGKYI GQLINMYAVT
 201 IGWWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL
 30 251 FAWNIWVAGS GLAVIFWFDG AMAELLKFAM IAAFVSAPVF AWLNYRLVKG
 301 DKRHRLTAGM NALAIVGLLY LAGFAVLFLL NLTGLLA*

ORF53ng-1 ([SEQ ID NO: 486](#)) and ORF53-1 ([SEQ ID NO: 480](#)) show 94.0% identity in 336 aa overlap:

35 orf53-1.pep 60 70 80 90 100 110
 ILTNLFKYPPFRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTA
 orf53ng-1 :|| |||||
 KKSCVYLWVFLILCIASATINAGAVAIVTA
 10 20 30

40 orf53-1.pep 120 130 140 150 160 170
 AIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIAM
 orf53ng-1 |||||
 AIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTLSIATLAAAGIAM
 40 50 60 70 80 90

45 orf53-1.pep 180 190 200 210 220 230
 SRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGI
 orf53ng-1 |||||
 SRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGI
 100 110 120 130 140 150

50 orf53-1.pep 240 250 260 270 280 290
 FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGWWSRPLVA

	orf53ng-1	FDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA 160 170 180 190 200 210
5		300 310 320 330 340 350
	orf53-1.pep	FIAFACMYGTTITTVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDDG : :
	orf53ng-1	FIAFACMYGTTITTVVDGYARAIAEPVRLLRGRDKTGNAELFAWNIWVAGSGLAVIFWFDDG 220 230 240 250 260 270
10		360 370 380 390 400 410
	orf53-1.pep	VMANLLKFAMIAAFVSAPVFPAWLNYRLVKGDDEKHKLTSGMNALALAGLIYLTGFTVLFLLL : : : : : : : : :
	orf53ng-1	AMAELLKFAMIAAFVSAPVFPAWLNYRLVKGDKRHRLTAGMNALAIVGLLYLAGFAVLFLLL 280 290 300 310 320 330
15		
	orf53-1.pep	NLAGMFKX : :
	orf53ng-1	NLTGLLAX

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 487>] (SEQ ID NO: 487):

25 1 ..TTGCGGAAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
 51 TGCGCTTGCC GGCTTGTTTT TTGTCCGCGC ACAATCCGAA CGCGAGTGGA
 101 TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG
 151 CTGCC TGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT
 201 GCTTTTCCAG GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTCTG
30 251 TCCGTTTCTG CCGAAACTAT CTGGCGCACG AATCCGAACC GGACAGGCC
 301 GTTCCGCT..

This corresponds to the amino acid sequence [[SEQ ID 488](#); [ORF58](#)] ([SEQ ID NO: 488](#); [ORF58](#)):

```

35      1  ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVS AW QEKKGEKQAE
      51  LPEIKDGM PD FPELALMLFH AVKTAVYWLF VGVRFCRNY LAHESEPD RP
     101  VPP..

```

Further work revealed the complete nucleotide sequence [<SEQ ID 489>] (SEQ ID NO: 489):

40 1 ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT

51 GTTTTTTTGTGTC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGT'TTGTCCG TGTCTGCTCCGT TTCTGCCGAA
5 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAACCGTG CCGATGTTC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGGAAGA AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTGC ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
10 501 AATTTTCGCCC GTCCGTCCGG TTTTAAAGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCGGA
651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG
15 751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACCGTC
851 ATGCAGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCCGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
20 1001 CGCGAATTTC TCGCCTGATT CCGGAAAGTC AGACGGTTGT CCGGAAACGG
1051 GATGTCGAAA TGCCGCTCTGA AACCGAAAAA GTTTTTCACGG AAACCGTTTC
1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAGTCCG GATATCCATA
1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC
25 1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAGGATTC GAGCAGGTGC
1301 AACGCAGCCG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTGAAT
1351 GGAGGTTGTC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCCGAAACC GAAGCGTTTCG
1451 GGCATGACAG TCAGCGGTTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
30 1501 CCGTCTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCCATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAAGTGTG
1651 GAAAACAGCA TCACCATCGA AGAAAAATG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGGCCCCGT AATTACGCGT TATGAAATCG
35 1751 AATCCGATGT CCGCGTGCGC GGCAATTCCG TTCTGAATCT GAAAAAGAT
1801 TTGGCGGCTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCTG AAACCATCCC
1851 CGGCAAAACC TGCATGGGTT TGGAACCTCC GAACCCGAAA CGCCAAATGA
1901 TACGCTGAG CGAAATCTTC AATTCGCCCC AGTTTGCCGA ATCCAAATCC
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGACAGC CCGTCCGAAC
40 2001 CGACTTGGGA AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGGTTCGG
2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
2101 GCGCCGGAAG ACGTGCATAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCCGTGC GTTACCGATA
2201 TGAAGCTGTC GGCAACCGC CTGAAGTGGT GTGTTAACGA AATGGAAAAA
45 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAATCTTG CGGGCTTCAA
2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT
2351 TCAGCCTCAC GCCCGACGAT CCCGAACCTT TGGAAAAACT GCCGTTTATC
2401 GTGGTCTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCCC GCCTCGCCA AAAAGCCCGC GCGGACGGCA
50 2501 TCCATTTGAT TCTTGCCACA CAACGCCCA GCGTCGATGT CATCACGGGT
2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
2601 AATCGACAGC CGCACGATTC TCGACCAAAT GGGCGCGGAA AACCTGCTCG
2651 GTCAGGGCGA TATGCTGTTC CTGCTGCCGG GTACTGCCTA TCCGACGCGC
2701 GTTACGCGC CGTTTGCTTC GGATGAAGAG GTGCACCGCG TGGTCAATA
55 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG
2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGACGAAACC
2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC
2901 CAGCATTTTC GCGGTACAGC GCGCCTTGCG TATCGGCTAC AACCGCGCCG
2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
60 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

This corresponds to the amino acid sequence [<SEQ ID 490; ORF58-1>] (SEQ ID NO: 490; ORF58-1):

```

1  MFWIVLIVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK
51 DGMPDFPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS
101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEEAADTED IATAVIDNRR
151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE
251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQGKG QAEAKSPDVS
301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR
351 DVEMPSETEN VFTETVSSVG YGGPVYDETA DIHIEEPAAP DAWVVEPPEV
401 PKVPMTAIDI QPPPPVSEIY NRTYEPPSGF EQVQRSRIAE TDHLADDVLN
451 GGWQEETAAI ADDGSEGAEE RSSGQYLSET EAFGHDSQAV CPFENVPSER
501 PSCRVSDETA DEGAFSEET GAVSEHLPTT DLLLPPLFNP EATQTEEELL
551 ENSITIEEKL AEFKVVKV VV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD
601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
651 KLTLALGQDI TGQPVVTDLG KAPHLVAGT TGS GSKSVGVN AMILSMLFKA
701 APEDVRMIMI DPKMLELSIY EGIPHLAPV VTD MKLAANA LNCV NEMEK
751 RYRLMSFMGV RNLAGFNQKI AEAAARGEKI GNPFSLT PDD PEPLKLPFI
801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLQGQDMLF LLPGTAYPQR
901 VHGAFADEE VHRVVEYLKQ FGEPDYVDDI LSGGGSEELP GIGRSGDDET
951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
1001 HNGNRTILVP LDNA*

```

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 (SEQ ID NO: 488) shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) (SEQ ID NO: 492) from strain A of *N. meningitidis*:

```

30      10      20      30      40      50      60
orf58.pep  LRETAYVLDSFD RYFVVALAGLFFVRAQSEREWMR EVSAWQEKKG EKQAELEPEIKDGMPD
           ::|||
orf58a      MFWIVLIVILLALAGLFFVRAQSEREWMR EVSAWQEKKG EKQAELEPEIKDGMPD
           10      20      30      40      50

35      70      80      90      100
orf58.pep  FP ELALMLFHAVKTA VYWLFGVVRF CRNYLAHESE PD RPVP
           |||
orf58a      FP ELALMLFHAVKTA VYWLFGVVRF CRNYLAHESE PD RPVP
           60      70      80      90      100      110

```

The complete length ORF58a nucleotide sequence [<SEQ ID 491>] (SEQ ID NO: 491) is:

```

1  ATGTTTTTGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51 GTTTTTTTGTC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGTTTGTCCG TGTCGTCCGT TTCTGCCGAA

```

5 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
 301 GCAAATCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
 351 AAACGGGACG GAAGAAGCGG AAACGGGAAGA AGCAGAAGCT GCGGAGGAAG
 401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
 451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
 501 AATTTTCGCCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
 551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
 601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCGA
 651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
 10 701 AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG
 751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
 801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCCGTC
 851 ATGCAGGGCA GGGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
 901 CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCCGCATG CCNGCCGCCG
 15 951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
 1001 CGCGGATTTT GCGCCTGATT CCGGAAAGTC GGACGGTTGT CGGGAACCGG
 1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAAT GTTTTCACGG AAANTGTTTC
 1101 GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAAACTGCC GATATCCATA
 1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
 20 1201 CCGAAAGTTC CCATGCCCGC AATNGATATT CCGCCGCCGC CTCCCGTATC
 1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GGCAGGATTC GAGCAGGTGC
 1301 AACGCAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTGAAT
 1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTAGGGG
 1401 TGTGGCAGAG CGGTCAAGCG GGCAATATTT GTCGGAACC GAAGCGTTCG
 25 1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
 1501 CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
 1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
 1601 TGCCGCCGCT GTTCAATCCC GGGGCGACGC AAACCGAAGA AGANCTGTTG
 1651 GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
 30 1701 CAAGGTTGTC GATTCTTATT CCGGCCCGGT GATTACGCGT TATGAAATCG
 1751 AACCCGATGT CGGCGTGCGC GGCAATTCG TTCTAAATCT GGAAAAAGAN
 1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCG AAACCATCCT
 1851 CGGCAAAACC TGTATGGGTT TGGAACCTCC GAACCCGAAA CGCCAAATGA
 1901 TACGCTGAG CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAAATCC
 35 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
 2001 CGACTTGGGC AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGGTTCCG
 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
 2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
 2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCCTGTC GTTACCGATA
 40 2201 TGAAGCTGGC GGCAAACGCG CTGAACTGGT GTGTTAACGA AATGGAAGAAA
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGTNTCAA
 2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAAATC GGCAACCCGT
 2351 TCAGCCTCAC GCCGACAAT CCCGAACCTT TGGANAAAT GCGGTTTATC
 2401 GTGGTCTGG TTTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
 45 2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
 2501 TCCATCTTAT CCTTGCCACA CAACGCCCA GTGTCGATGT CATCAGGGT
 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
 2601 AATCGACAGC CGCACGATTC TTGACCAAAT GGGTGCGGAA AACCTGCTCG
 2651 GGCAGGGCGA TATGCTGTTT CTGCCGCCGG GTACGGCCTA TCCGCAAGCG
 50 2701 GTTCACGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA
 2751 TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
 2801 GTATGTCCGA CGATTTGCTG GGAATCAGCC GGAGCGGCGA CGGCGAAACC
 2851 GATCCGATGT ACGACGAGGC CGTGTGNGTT GTTTTGAAAA CGCGCAAAGC
 2901 CAGCATTTCT GCGGTGCAGC GCGCATTGCG TATCGGCTAT AATCGCGCCG
 55 2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

This encodes a protein having amino acid sequence [<SEQ ID 492>] (SEQ ID NO: 492):

1 MFWIVLVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK

```

5      51  DGMPDFPELA  LMLFHAVKTA  VYWLFGVVVR  FCRNYLAHES  EPDRPVPPAS
      101  ANRADVPTAS  DGYSDSGNGT  EEAETEEAEA  AEEEAADTED  IATAVIDNRR
      151  IPFDRSIAEG  LMPSESEISP  VRPVFKEITL  EEATRALNSA  ALRETKKRYI
      201  DAFEKNETAV  PKVRVSDTPM  EGLQIIGLDD  PVLQRTYSRM  FDADKEAFSE
      251  SADYGFEPYF  EKQHPSAFSA  VKAENARNAP  FRRHAGQGKG  QAEAKSPDVS
      301  QGQSVSDGTA  VRDAXRRVSV  NLKEPNKATV  SAEARISRLI  PESRTVVGKR
      351  DVEMPSETEN  VFTEXVSSVG  YGXPVYDETA  DIHIEEPAAP  WDAWVVEPPEV
      401  PKVPMAXDI  PPPPVSEIY  NRTYEPPAGF  EQVQRSRIAE  TDHLADDVLN
      451  GGWQEETAAI  ANDGSEGVAE  RSSGQYLSET  EAFGHDSQAV  CPFENVPSER
      501  PSRRAXDTEA  DEGAFQSEET  GAVSEHLPTT  DLLLPPLFNP  GATQTEEXLL
      551  XNSITIEEKX  AEFKVKVKVV  DSYSGPVITR  YEIEPDVGVR  GNSVLNLEKX
      601  LARSLGVASI  RVVETILGKT  CMGLELPNPK  RQMIRLSEIF  NSPEFAESKS
      651  KLTLALGQDI  TGQPVVTDLG  KAPHLLVAGT  TSGGKSVGVN  AMILSMLFKA
      701  APEDVRMIMI  DPKMLELSIY  EGIPHLLAPV  VTDMKLAANA  LNWCVNEMEK
      751  RYRLMSFMGV  RNLAGXNQKI  AEAARGEKI  GNPFSLTDPN  PEPLXKLFFI
      801  VVVDEFADL  MMTAGKKIEE  LIARLAQKAR  AAGIHLILAT  QRPSVDVITG
      851  LIKANIPTRI  AFQVSSKIDS  RTILDQMGAE  NLLQGQDMLF  LPPGTAYPQR
      901  VHGAFAASDEE  VHRVVEYLKQ  FGEPDYVDDX  LSGGMSDDL  GISRSGDGET
      951  DPMYDEAVSV  VLKTRKASIS  GVQRALRIGY  NRAARLIDQM  EAEGIVSAPE
      1001 HNGNRTILVP  XDNA*

```

ORF58a (SEQ ID NO: 492) and ORF58-1 (SEQ ID NO: 490) show 96.6% identity in 1014 aa overlap:

```

25      10      20      30      40      50      60
      orf58a.pep  MFWIVLIVILLALLAGLFFVRAQSEREWREVSAAWQEKKGKQELPEIKDGMDFPELA
      orf58-1     MFWIVLIVILLALLAGLFFVRAQSEREWREVSAAWQEKKGKQELPEIKDGMDFPELA
                10      20      30      40      50      60

30      70      80      90      100     110     120
      orf58a.pep  LMLFHAVKTAVYWLFGVVRFRCRNYLAHESEPDRPVPPASANRADVPTASDGYSDSGNGT
      orf58-1     LMLFHAVKTAVYWLFGVVRFRCRNYLAHESEPDRPVPPASANRADVPTASDGYSDSGNGT
                70      80      90      100     110     120

35      130     140     150     160     170     180
      orf58a.pep  EEAETEEAEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPVRPVFKEITL
      orf58-1     EEAETEEAEAAEEEAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPVRPVFKEITL
                130     140     150     160     170     180

40      190     200     210     220     230     240
      orf58a.pep  EEATRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSRM
      orf58-1     EEATRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSHM
                190     200     210     220     230     240

45      250     260     270     280     290     300
      orf58a.pep  FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS
      orf58-1     FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS
                250     260     270     280     290     300

50      310     320     330     340     350     360
      orf58a.pep  QGQSVSDGTAVRDAXRRVSVNLKEPNKATVSAEARISRLIPESRTVVGKRDEVMPSETEN
      orf58-1     QGQSVSDGTAVRDARRRVSVNLKEPNKATVSAEARISRLIPESQTVVGKRDEVMPSETEN

```

-374-

		310	320	330	340	350	360
	orf58a.pep	370	380	390	400	410	420
	orf58-1	370	380	390	400	410	420
5	orf58-1	370	380	390	400	410	420
	orf58a.pep	430	440	450	460	470	480
	orf58-1	430	440	450	460	470	480
10	orf58-1	430	440	450	460	470	480
	orf58a.pep	490	500	510	520	530	540
	orf58-1	490	500	510	520	530	540
15	orf58-1	490	500	510	520	530	540
	orf58a.pep	550	560	570	580	590	600
	orf58-1	550	560	570	580	590	600
20	orf58-1	550	560	570	580	590	600
	orf58a.pep	610	620	630	640	650	660
	orf58-1	610	620	630	640	650	660
25	orf58-1	610	620	630	640	650	660
	orf58a.pep	670	680	690	700	710	720
	orf58-1	670	680	690	700	710	720
30	orf58-1	670	680	690	700	710	720
	orf58a.pep	730	740	750	760	770	780
	orf58-1	730	740	750	760	770	780
35	orf58-1	730	740	750	760	770	780
	orf58a.pep	790	800	810	820	830	840
	orf58-1	790	800	810	820	830	840
40	orf58-1	790	800	810	820	830	840
	orf58a.pep	850	860	870	880	890	900
	orf58-1	850	860	870	880	890	900
45	orf58-1	850	860	870	880	890	900
	orf58a.pep	910	920	930	940	950	960
	orf58-1	910	920	930	940	950	960

	orf58-1	VHGA	FASD	EDEV	HVRV	VEYL	KQFG	EPDY	VDDI	LSGG	GSEEL	PGRS	GDDE	TDPY	DEAV	SV	
		910			920			930			940			950		960	
		970			980			990			1000			1010			
5	orf58a.pep	VLKTR	KASIS	GVQR	ALRIG	YNRA	ARLI	DQME	AEGIV	SAPEN	HNGNR	TILVP	XDNAX				
	orf58-1	VLKTR	KASIS	GVQR	ALRIG	YNRA	ARLI	DQME	AEGIV	SAPEN	HNGNR	TILV	PLDNAX				
		970			980			990			1000			1010			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 (SEQ ID NO: 488) shows complete identity over a 9aa overlap with a predicted ORF
10 (ORF58ng) (SEQ ID NO: 494) from *N. gonorrhoeae*:

orf58.pep	ALMLFHAVKTAVYWLFGVVRFRCRNYLAHESEPD	103
orf58nq	SEPDRPVPPASANRADVPTASDGYSDSGNG	30

15 The ORF58ng nucleotide sequence [<SEQ ID 493>] (SEQ ID NO: 493) is predicted to encode a protein having partial amino acid sequence [<SEQ ID 494>] (SEQ ID NO: 494):

	1	SEPDRPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAEETAAE	AAEEEAADTE
	51	DIATAVIDNR	RIPFDRSIAE	GLMQSESKTS	PVRPVFKEIT	LEEATRALSS
20	101	AALRETKKRY	IDAFEKNGTA	VPKVRVSDTP	MEGLQIIGLD	DPVLQRTYSR
	151	MFDADKEAFS	ESADYGFEFY	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK
	201	GQAEAKSPDV	SQGQSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL
	251	IPESRTVVVGK	RDVEMPSETE	NVFTETVSSV	GYGGPVYDEA	ADIHIEEPAA
	301	PDAWVVEPPE	VPEVAVPEID	ILPPPPVSEI	YNRTYEPPAG	FEQAQRSRIA
25	351	ETDHLAADVL	NGGWQEETAA	IADDGSEGAA	ERSSGQYLSE	TEAFGHDSQA
	401	VCPFEDVPSE	RPSCRVSDEE	ADEGAFQSEE	TGAVSEHLPT	TDLLLPPLFN
	451	PEATQTEEEL	LENSITIEEK	LAEFKVKVKV	VDSYSGPVIT	RYEIEPDVGV
	501	RGNSVLNLEK	DLARSLGVAS	IRVVETIPGK	TCMGLLEPNP	KRQMIRLSEI
	551	FNSPEFAESK	SKLTLALGQD	ITGQPVVTDL	GKAPHLLVAG	TTGSGKSGVG
	601	NAMILSMLFK	AAPEDVRMIM	IDPKMLELSI	YEGITHLLAP	VVTDMKLAAN
30	651	ALNWCVNEME	KRYRLMSFMG	VRNLAGFNQK	IAEAAARGEK	IGNPFSLTPD
	701	DPEPLEKLPF	IVVVVDEFAD	LMMTAGKKIE	ELIARLAQKA	RAAGIHLILA
	751	TQRPSVDVIT	GLIKANIPTR	IAFQVSSKID	SRTILDQMG	ENLLQGQDML
	801	FLPPGTAYPO	RVHGAFASDE	EVHRVVEYLK	QFGEPTYVDD	ILSGGGSEEL
	851	PGIGRSGDGE	TDPMYDEAVS	VVLKTRKASI	SGVQRALRIG	YNRAARLIDQ
35	901	MEAEGRIVSAP	EHNGNRTILV	PLDNA*		

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng (SEQ ID NO: 494) and FtsK (accession number p46889) (SEQ ID NO: 1142) show a 65 % amino acid identity in 459 overlap:

ORF58ng: 467 IEKLAIEFKVKVKVVDSSYSGPVITRYEIEPDVGVGRGNSVLNLEKDLARSLGVASIRVVET 526
+E +LA+F++K VV+ GPVITR+E+ GV+ +NL +DLARSL ++RVVE
FtsK: 868 VEARLADFRIKADVVNYSPPGPVITRFELNLAGVKAARISNLSRDLARSLSTVAVRVVEV 927

	ORF58ng:	527	IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL	586
	FtsK:	928	IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL	
5	ORF58ng:	587	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMDK	646
	FtsK:	988	LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK	
	ORF58ng:	647	LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAAARGEKIGNPFSLTPDDPEP--	704
	FtsK:	1048	AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +	
10	ORF58ng:	705	--LEKLPFIVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL	762
	FtsK:	1108	L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL	
15	ORF58ng:	763	IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQVRVHGAFASDEEV	822
	FtsK:	1168	IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV	
	ORF58ng:	823	HRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG	882
	FtsK:	1228	H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG	
20	ORF58ng:	883	VQRALRIGYNRAARLIDQMEAEIVSAPEHNGNRTILVP	921
	FtsK:	1287	VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P	

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be [[SEQ ID 495](#)] ([SEQ ID NO: 495](#)):

25	1	ATGTTTTGGA	TAGTTTTGAT	CGTTATtgtg	TTGCTTGCGC	TTGCCGGCCT
	51	GTTTTTTGTC	CGCGCACAAAT	CCGAACGCGA	GTGGATGCGC	GAGGTTTCTG
	101	CGTGGCAGGA	AAAGAAAGGG	GAAAAACAGG	CGGAGCTGCC	TGAAATCAAA
	151	GACGGTATGC	CCGATTTTCC	CGAGTTTTC	CTGATGCTTT	TCCATGCCGT
	201	CAAAACGCGA	GTGTATTGGC	TGTTTGTCGG	TGTCGTCCGT	TTCTGCCGAA
	251	ACTATCTGGC	GCACGAATCC	GAACCGGACA	GGCCCGTTCC	GCCTGCTTCT
30	301	GCAAACCGTG	CGGATGTTCC	GACCGCATCC	GACGGGTATT	CAGACAGTGG
	351	AAACGGGACG	GAAGAAGCGG	AAACGGAAGC	AGCAGAAGCT	GCGGAGGAAG
	401	AGGCTGCCgA	TACgGAAGAC	ATTGCAACTG	CCGTAATCGA	CAACCGCCGC
	451	ATCCcatTCG	ACCGGAGTAT	TGCTGAAGGG	TTGATGCAGT	CTGAAAGCAA
35	501	AACTTCGCCC	GTCCGTCCGG	TTTTTAAGGA	AATCACTTTG	GAAGAAGCAA
	551	CGCGTGCTTT	AAGCAGCGCG	GCTTTAAGGG	AAACGAAAAA	ACGCTATATC
	601	GATGCATTTG	AGAAAAACGG	AACAGCCGTC	CCCAAAGTAC	GCGTGTCGGA
	651	TACCCCGATG	GAAGGGCTGC	AGATTATCGG	TTTGACGAC	CCTGTGCTTC
	701	AACGCACGTA	TTCCCGTATG	TTTGATGCGG	ACAAAGAAGC	GTTTTCCGAG
	751	TCTGCGGATT	ACGGATTGGA	GCCGTATTTT	GAGAAGCAGC	ATCCGTCTGC
40	801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCACG	GAATGCGCCG	TTCCGCCGTC
	851	ATGCAGGGCA	GGAGAAAGGG	CAGGCGGAGG	CAAAATCCCC	GGATGTTTCC
	901	CAAGGGCAGT	CCGTTTCAGA	CGGCACAGCC	GTCCGCGATG	CCCGCCGCCG
	951	CGTTTCCGTC	AATTTGAAAG	AACCGAACAA	GGCAACGGTT	TCTGCGGAGG
45	1001	CGCGGATTTC	GCGCCTGATT	CCGGAAAGTC	GGACGTTGT	CGGAAACGG
	1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAAAT	GTTTTACGG	AAACCGTTTC
	1101	GTCTGTGGGA	TACGGCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
	1151	TTGAAGAGCC	TGCCGCGCCC	GATGCTTGGG	TGGTCGAACC	ACCGAAGTG
	1201	CCGGAGGTAG	CCGTACCGCA	AATCGATATT	CTGCCGCCGC	CTCCGCTATC
	1251	GGAAATCTAC	AACCGTACCT	ATGAGCCGCC	GGCAGGATTC	GAGCAGGCGC
50	1301	AACGCAGCCG	CATTGCCGAA	ACCGACCATC	TTGCCGCTGA	TGTTTTGAAT
	1351	GGAGGTTGGC	AGGAGGAAAC	CGCCGCTATT	GCAGATGACG	GCAGTGAGGG

-377-

5 1401 TGC GGCAGAG CCGTCAAGCG GGCAATATCT GTCGGAACCC GAAGCGTTTCG
 1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAGATGTGCC GTCTGAACGC
 1501 CCGTCCCTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
 1551 GGAAGAGACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
 1601 TGCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG
 1651 GAAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
 1701 CAAGGTTGTC GATTCTTATT CCGGCCCCGT GATTACGCGT TATGAAATCG
 1751 AACCCGATGT CGGCGTGCGC GGCAATTCCG TTCTGAATTT GGAAAAAGAC
 10 1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTGC AAACCATCCC
 1851 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA
 1901 TACGCCTGAG CGAAATTTTC AATTCGCCCC AGTTTGCCGA ATCCAAATCC
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATT ACCGGACAGC CCGTCGTAAC
 2001 CGACTTGGGC AAAGCACCGC ATTTGCTGGT TGCCCGCACG ACCGTTTCGG
 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
 15 2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
 2151 GAGCATTTAC GAAGGCATCA CGCACCTGCT CGCCCTGTGC GTTACCGATA
 2201 TGAAGCTGGC GGCAAAACGC CTGAACTGGT GTGTTAACGA AATGGAAGAA
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGCTTCAA
 2301 CCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT
 20 2351 TCAGCCTCAC GCCCGACGAT CCCGAACCTT TGGAAAAACT GCCGTTTATC
 2401 GTGGTCTGG TCGATGAGTT TGCCGATTTG ATGATGACGG CAGGCAAGAA
 2451 AATCGAAGAA CTGATTGCGC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
 2501 TCCACCTTAT CCTTGCCACA CAACGCCCA GCGTCGATGT CATCAGGGGT
 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
 25 2601 AATCGACAGC CGCACGATTC TCGACCAAAT GGGCGCGGAA AACCTGCTCG
 2651 GTCAGGGCGA TATGCTGTTC CTGCCGCCGG GTACTGCCTA TCCGACGCGC
 2701 GTTCACGGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTGAATA
 2751 TCTGAAGCAG TTTGGCGAGC CGGACTATGT TGACGATATT TTGAGCGGCG
 2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGGCGAAACC
 30 2851 GATCCGATGT ACGACGAGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC
 2901 CAGCATTTTCG GGCCTACAGC GCGCCTTGCG CATCGGCTAC AACCGCGCCG
 2951 CGCGTCTGAT TGACCAAATG GAAGCGGAAG GCATTGTGTC CGCACCAGAA
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

35 This corresponds to the amino acid sequence [<SEQ ID 496; ORF58ng-1>] (SEQ ID NO: 496;
ORF58ng-1):

1 MFWIVLIVIV LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK
 51 DGMPDFPEFS LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS
 40 101 ANRADVPTAS DGYSDSNGNT EEAETEAAEA AEEEEADTED IATAVIDNRR
 151 IPFDRSIAEG LMQSEKSTSP VRPVFKEITL EEATRALSSA ALRETKKRYI
 201 DAFEKNGTAV PKVRVSDTPM EGLQIIIGLDD PVLQRTYSRM FDADKEAFSE
 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQEKQ QAEAKSPDVS
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
 351 DVEMPSETEN VFTETVSSVG YGGPVYDEAA DIHIEEPAAP DAWVVEPPEV
 45 401 PEVAVPEIDI LPPPPVSEIY NRTYEPPAGF EQAQRSRIAE TDHLAADVLN
 451 GGWQEETAAI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFEDVPSEK
 501 PSRVSDETA DEGAFAQSEET GAVSEHLPTT DLLLPLPLFNP EATQTEEELL
 551 ENSITIEEKL AEFKVKVKVV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 50 651 KLTLALGQDI TGQPVVTDLG KAPHLLVAGT TGSGKSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGITHLLAPV VTDMLAANA LNWCVNEMEK
 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTPDD PEPELEKLPFI
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LPPGTAYPQR
 55 901 VHGAFADEE VHRVVEYLKQ FGEPDYVDDI LSGGGSEELP GIGRSGDGET
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
 1001 HNGNRTILVP LDNA*

ORF58ng-1 (SEQ ID NO: 496) and ORF58-1 (SEQ ID NO: 490) show 97.2% identity in 1014 aa overlap:

5	orf58-1.pep	10 20 30 40 50 60	MFWIVLIVILLALAGLFFVRAQSEREWREVS	AWQEKKGKQAE	LPEIKDGM	PDFPELA
	orf58ng-1	10 20 30 40 50 60	MFWIVLIVIVLLALAGLFFVRAQSEREWREVS	AWQEKKGKQAE	LPEIKDGM	PDFPEFS
10	orf58-1.pep	70 80 90 100 110 120	LMLFHAVKTAVYWL	FVGVVRCRNYLA	HESEPD	RPVPPASANRADVPTASDGYSDSGNGT
	orf58ng-1	70 80 90 100 110 120	LMLFHAVKTAVYWL	FVGVVRCRNYLA	HESEPD	RPVPPASANRADVPTASDGYSDSGNGT
15	orf58-1.pep	130 140 150 160 170 180	EEAE	TEEA	EAEEEEAA	ADTEDIATAVIDNRRIPFDRSIAEGLMPSESEIS
	orf58ng-1	130 140 150 160 170 180	EEAE	TEEA	EAEEEEAA	ADTEDIATAVIDNRRIPFDRSIAEGLMQSESKTSPVRPVFKEITL
20	orf58-1.pep	190 200 210 220 230 240	EEATRALNSAAL	RETKKRYIDAF	EKNETAVPKVR	VS
	orf58ng-1	190 200 210 220 230 240	EEATRALSSAAL	RETKKRYIDAF	EKNGTAVPKVR	VS
25	orf58-1.pep	250 260 270 280 290 300	FDADKEAF	SESADYGF	EPYFEKQHP	SAFSAVKAENARNAPFHRHAGQGKGQAEAKSPDVS
	orf58ng-1	250 260 270 280 290 300	FDADKEAF	SESADYGF	EPYFEKQHP	SAFSAVKAENARNAPFRRHAGQKQAEAKSPDVS
30	orf58-1.pep	310 320 330 340 350 360	QQQSVSDGTAV	RDARRRVSVNL	KEPNKATVSA	EARISRLIPESQTVVGKRDVEMPSETEN
	orf58ng-1	310 320 330 340 350 360	QQQSVSDGTAV	RDARRRVSVNL	KEPNKATVSA	EARISRLIPESRTVVGKRDVEMPSETEN
35	orf58-1.pep	370 380 390 400 410 420	VFTETVSSVG	YGGPVYDE	TADIHIEEPA	APDAWVVEPPEVPKVPMTAIDIQPPPPVSEIY
	orf58ng-1	370 380 390 400 410 420	VFTETVSSVG	YGGPVYDE	AADIHIEEPA	APDAWVVEPPEVAVPEIDILPPPPVSEIY
40	orf58-1.pep	430 440 450 460 470 480	NRTYEPPSG	FQVQSR	IAETDHLADD	VLNGGWQEETA
	orf58ng-1	430 440 450 460 470 480	NRTYEPPAG	FQVQSR	IAETDHLAAD	VLNGGWQEETA
45	orf58-1.pep	490 500 510 520 530 540	EAFGHDSQAV	CPFENVP	SERPSCR	VSDTEADEGAF
	orf58ng-1	490 500 510 520 530 540	EAFGHDSQAV	CPFEDVP	SERPSCR	VSDTEADEGAF

5	orf58-1.pep	550	560	570	580	590	600
	orf58ng-1	EATQTEEELENSITIEEKLAEFKVKVKVDSYSGPVITRYEIEPDVGVRGNSVLNLEKD					
10	orf58-1.pep	610	620	630	640	650	660
	orf58ng-1	LARSLGVASIRVVETIPGKTCMGLLEPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
15	orf58-1.pep	670	680	690	700	710	720
	orf58ng-1	TGQPVVTDLGKAPHLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
20	orf58-1.pep	730	740	750	760	770	780
	orf58ng-1	EGIPHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
25	orf58-1.pep	790	800	810	820	830	840
	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
30	orf58-1.pep	850	860	870	880	890	900
	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMFLPPGTAYPQR					
35	orf58-1.pep	910	920	930	940	950	960
	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
40	orf58-1.pep	970	980	990	1000	1010	
	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					

Furthermore, ORF58ng-1 (SEQ ID NO: 496) shows significant homology to the *E.coli* protein FtsK (SEQ ID NO: 1142):

sp|P46889|FTSK_ECOLI CELL DIVISION PROTEIN FTSK)gi|1651412|gnl|PID|d1015290 (D1 division protein FtsK [Escherichia coli])gi|1651418|gnl|PID|d1015296 (D90727) Cell division protein FtsK [Escherichia coli])gi|1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329
Score = 576 bits (1469), Expect = e-163
Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)

Query: 556 IEEKLAEFKVKVKVDSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 615
 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE
 Sbjet: 868 VEARLADFRIKADVNYSPGPVITRFELNLAGVKAARISNLSRDLARSLSTVAVRVVEV 927

Query: 616 IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 675
 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL
 Sbjet: 928 IPGKPYVGLELPNKKRQTVYLREVLDNAKFRDNPSPLTVVLGKDIAEPVVADLAKMPHL 987

Query: 676 LVAGTTGSGKSVGVMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDK 735
 LVAGTTGSGKSVGVMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDK
 Sbjet: 988 LVAGTTGSGKSVGVMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLTEVVTDK 1047

Query: 736 LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFLTPDDPEP-- 793
 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +
 Sbjet: 1048 DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPPIPDYKPGDSMDAQH 1107

Query: 794 --LEKLPIFIVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSPVDVITGL 851
 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSPVDVITGL
 Sbjet: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSPVDVITGL 1167

Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQDMLFLPPGTAYPQRVHGAFASDEEV 911
 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV
 Sbjet: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGMMLYSGPNSTLPVRVHGAFVRDQEV 1227

Query: 912 HRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 971
 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG
 Sbjet: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286

Query: 972 VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP 1010
 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P
 Sbjet: 1287 VQRQFRIGYNRAARIIEQMEAQGIVSEQHNGNREVLAP 1325

25 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 497>] (SEQ ID NO: 497):

30 1 ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
 51 CATTTTCGTC GTCCTCTTGG CCGTATTGGT CTCCACGCAG GCAATCAACC
 101 TGCTCGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTTG
 151 GCATTGGTCG GCTTCTGGGT C.....
 //
 35 901A TTGCCATCGG TTTGTTTTTA ATTTACCAA ACGGGCTGAC
 951 CCTGCTTTT GAAGCCGTGG AAGACGCAA AATCCATTTT TGGCTCGGAC
 1001 TGCTGCCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTTGCGC
 1051 GTCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAGTCT
 40 1101 GACATTGAAA GCGGAAAT GA

This corresponds to the amino acid sequence [<SEQ ID 498; ORF101>] (SEQ ID NO: 498; ORF101):

```

1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
51  ALVGFVW... ..
5  301  ...IAIGLFL IYQNLTLFL EAVEDGKIHF WLGLLPMHII MFVLALILLR
351  VRSMPSQPFV QAVGKSLTLK GGK*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 499>] (SEQ ID NO: 499):

```

10  1  ATGATTTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
    51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
101  TGCTCGGCCG TGCCGCCGAC GGGCGTGTCT CCATCGATGC CGTGTTGGCA
151  TTGGTCGGCT TCTGGGTCAT CGGTATGACG CCGCTTTTGC TGGTGTTGAC
201  CGCATTTATC AGTACGTTGA CCGTGTTGAC CCGCTACTGG CGCGACAGCG
251  AAATGTCGGT CTGGCTATCC TGC GGATTGG CATTGAAACA ATGGATACGC
15  301  CCGGTGATGC AGTTTGCCGT GCCGTTTGCC GTTTTGGTTG CCGTCATGCA
351  GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
401  TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
451  AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTCGAAA CCTTCGATAC
501  CGAATCCGGC ATCATGAAAA ACCTGTTTCT GCGCGAACAG GACAAAAACG
20  551  GCGGCGACAA CATCATCTTC GCCAAAGAAG GTAACCTTCT GCTGAACGAC
601  AACAAACGCA CGTCTGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
651  CGGACGCGCC GACTACAATC AGGTTTCCTT CCAAAAACTC AACCTGATTA
701  TCAGACCCAC GCCCAAACCT ATCGACCCCG TTTCCACCG CCGTACCATT
751  CCGACCGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
25  801  GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTGCTTG
851  CCGTGCCGCT TTCCTATTTT AACC CGCGCA GCGGACATAC CTACAATATC
901  TTGATTGCCA TCGGTTTGTT TTTAATTAC CAAAACGGGC TGACCCTGCT
951  TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGA CTGCTGC
30  1001 CTATGCACAT TATCATGTTT GCCGTTGCAC TCATCCTGTT GCGCGTCCGC
1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGCGCGA AAATGA

```

This corresponds to the amino acid sequence [<SEQ ID 500; ORF101-1>] (SEQ ID NO: 500; ORF101-1):

```

35  1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
    51  LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101  PVMQFAVPFA VLAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
151  SLGKRNGRVY FVETFDTEG IMKNLFLREQ DKNGGDN IIF AKEGNFSLND
40  201  NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
251  PTAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301  LIAIGLFLIY QNGLTLLFEA VEDGKIHF WL GLLPMH IIMF AVALILLVR
351  SMPSQPFWQA VGKSLTLKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

45 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 (SEQ ID NO: 498) shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) (SEQ ID NO: 502) from strain A of *N. meningitidis*:

		10	20	30	40	50
orf101.pep		MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX				
5	orf101a	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRX-AIDAVLALVGFVXXM				
		10	20	30	40	50
		//				
				90	100	110
10	orf101.pepIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL				
	orf101a	LTVSVLLLCLLAVPLSYFNPRSGHTYNILXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL				
		280	290	300	310	320 330
		120	130	140	150	
15	orf101.pep	LPMHIIMFVLALILLRVRSMPSQPFQAVGKSLTLKGGKX				
	orf101a	LPMHIIMFVIAIVLLRVRSMPSQPFQAVGKSLTLKGGKX				
		340	350	360	370	

The complete length ORF101a nucleotide sequence [<SEQ ID 501>] (SEQ ID NO: 501) is:

20	1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
	51	CATTTTCGTC	GTCTCTTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
	101	TGCTCGGCCN	TGCCGCCGAC	NGCGTNTCG	CCATCGATGC	CGTGTGGCA
	151	TTGGTCGGCT	TCTGGGTCNN	NNGNATGACG	CCGCTTTTGC	TNGTGTGAC
25	201	CGCATTTATC	AGTACGTTGA	CCGTGTTGAC	CCGCTACTGG	CGNGACAGCG
	251	AAATGTCGGT	CTGNTATCC	TGCGGATTGG	CATTGAAACA	ATGGATACGC
	301	CCGGTGATGC	AGTTTGCCGT	GCCGTTTGCC	GTTTGGTTG	CCGTCATGCA
	351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTACG	CAGCCGCGAA	TACGCTGAAA
	401	TCCTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAGGCAGG	CGGTTCAAC
	451	AGTTTGGGCA	AGCGCAACGG	CAGGGTTTAT	TTGTGCGAAA	CCTTCGATAC
30	501	CGAATCCGGC	ATCATGAAAA	ACCTGTTCTT	GCGCGAACAG	GACAAAAACG
	551	GCGGCGACAA	CATCATCTTC	NCCAAAGAAA	GTAACCTTCT	GCTGAACGAC
	601	AACAAACGCA	CGCTCGAATT	GCGCCACGGC	TACCGTTACA	GCGGCACGCC
	651	CGGACGCGCC	GACTACAATC	AGGTTTCCTT	CCNAAAATC	AACCTGATTA
35	701	TCAGCACCAC	GCCCAAATC	ATCGACCCCG	TTCCCACCG	CCGTACNATN
	751	CCNACNGCCC	AACCTGATTG	CAGCAGCAAC	CCGCAACATC	ANGCGGAATT
	801	GATGTGGCGC	ATCTCGCTGA	CCGTCAGCGT	CCTCCTACTC	TGCCTGCTTG
	851	CCGTGCCGCT	TTCTATTTC	AACCCGCGCA	GCGGACATAC	CTACAATATC
	901	TTGANTGCCA	TCGTTTGTG	TTTAATTAC	CAAAACGGGC	TGACCCTGCT
40	951	TTTTGAAGCC	GTGGAAGACG	GCAAAATCCA	TTTTTGGCTC	GGACTGCTGC
	1001	CTATGCACAT	CATCATGTTT	GTCATCGCAA	TCGTACTTCT	GCGCGTCCGC
	1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGGCG	GTTGGCAAAA	GTCTGACATT
	1101	GAAAGCGGGA	AAATGA			

This encodes a protein having amino acid sequence [<SEQ ID 502>] (SEQ ID NO: 502):

45	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGXAAD	XRXAIDAVLA
	51	LVGFVXXMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWXS	CGLALKQWIR
	101	PVMQFAVPFA	VLVAVMQLWV	IPWAEIRSRE	YAEILKQKQE	LSLVEAGGFN
	151	SLGKRNGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	XKESNFSLND
	201	NKRTLRLRHG	YRYSGTPGRA	DYNQVSFXKL	NLIISTTPKL	IDPVSHRRTX
50	251	PTAQLIGSSN	PQHXAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
	301	LXAIGLFLIY	QNGLTLLFEA	VEDGKIHFWL	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSQPFWQA	VGKSLTLKGG	K*		

ORF101a (SEQ ID NO: 502) and ORF101-1 (SEQ ID NO: 500) show 95.4% identity in 371 aa overlap:

```

5      orf101a.pep  MIYQRNLIKELSFTAVGIFVVLLAVLVSTQAINLLGXAADXRXAIDAVLALVGFVWXXMT  60
      orf101-1     MIYQRNLIKELSFTAVGIFVVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWVIGMT  60

      orf101a.pep  PLLLVLTAFISTLTVLTRYWRDSEMSVWXSCLGALKQWIRPVMQFAVPFAVLVAVMQLWV  120
      orf101-1     PLLLVLTAFISTLTVLTRYWRDSEMSVWLSCLGALKQWIRPVMQFAVPFAVLVAVMQLWV  120

10     orf101a.pep  IPWAELRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ  180
      orf101-1     IPWAELRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ  180

      orf101a.pep  DKNGGDNIIIFXKESNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFXKLNLIISTTPKL  240
      orf101-1     DKNGGDNIIIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNLIISTTPKL  240

15     orf101a.pep  IDPVSHRRTXPTAQLIGSSNPQHXAELMWIRISLTVSVLLLCLLAVPLSYFNPRSGHTYNI  300
      orf101-1     IDPVSHRRTIPTAQLIGSSNPQHXAELMWIRISLTVSVLLLCLLAVPLSYFNPRSGHTYNI  300

      orf101a.pep  LXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA  360
20     orf101-1     LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFAVALILLRVRSMPSQPFWQA  360

      orf101a.pep  VGKSLTLKGGK  371
      orf101-1     VGKSLTLKGGK  371

```

Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF101 (SEQ ID NO: 498) shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) (SEQ ID NO: 504) from *N. gonorrhoeae*:

```

30     orf101.pep  MIYQRNLIKELSFTAVGIFVVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVW  57
      orf101ng     MIYQRNLIKELSFTAVGIFVVLLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFVWIGM  59

                                     //

      orf101.pep                                     IAIGLFLIYQNGLTLLFEAVEDGKIHFWLG  333
      orf101ng     SLTVSVLLLCLLAVPLSYFNPRSGHTYNIILIAIGLFLIYQNGLTLLFEAVEDGKIHFWLG  331

35     orf101.pep  LLPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGK  373
      orf101ng     LLPMHIIMFVIAIVLLRVRSMPSQPFWQAVG  362

```

The ORF101ng nucleotide sequence [<SEQ ID 503>] (SEQ ID NO: 503) is predicted to encode a protein having partial amino acid sequence [<SEQ ID 504>] (SEQ ID NO: 504):

```

1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
5  51  LVGFWVIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101 PVMQFAVPFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLKD
201 NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
251 STAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301 LIAIGLFLIY QNGLTLLFEA VEDGKIHF L GLLPMHIIMF VIAIVLLRVR
10 351 SMPSQPFWQA VG...

```

Further work revealed the complete nucleotide sequence [<SEQ ID 505>] (SEQ ID NO: 505):

```

1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
15 51  CATTTTCGTC GTCTCTTGG CGTGTTGGT GTCCACGCAG GCGATCAACC
101 TGCTTGCCG CGCAGCTGAC GGGCGTGTG CCATCGATGC CGTGTTGGCC
151 TTAGTCGGCT TCTGGGTCAT CGGTATGACC CCGCTTTTGC TGGTGTGAC
201 CGCATTCATC AGCACGCTGA CCGTATTGAC CCGCTACTGG CCGACAGCG
251 AAATGTCGGT CTGGCTATCC TGGCGATTGG CGTTGAAACA GTGGATACGC
301 CCCGTCATGC AGTTTGCCGT GCCGTTTGCC ATCCTGATTG CCGTCATGCA
20 351 GCTTTGGGTG ATACCGTGGG CAGAGCTGCG CAGCCGCGAA TATGCCGAAA
401 TTTTGAAGCA GAAGCAGGAA TTGTCTTTGG TGAAGCCGG CGAGTTCAAT
451 AACTTGGGCA AGCGCAACGG CAggggtttaT Ttcgtcgaaa CCTTTGACAC
501 CGaatccgGC ATCATGAAAA ACCTGTtcct GcGCGAACAG GACAAAAACG
25 551 gcgcgacaaA CATCATCTTC GCcaaaGAag gtaactTctc gctgaaggac
601 AACAAAcgca cgctcgaaATT GCGCCACGGC TACCGTTACA GCGGcacgcC
651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAacTc aacctgAtta
701 TCAGCACCAC GCCCAAacTT ATCGaccCCG TTCCCAACCG CCGCACCATT
751 tcgacCGCCC AAcTGATTGG CAGCAGCAAT CCGCAACATC AGGCAGAATT
30 801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG
851 CCGTGCCGCT TTCCTATTTT AACCCGCGCA GCGGACATAC CTACAATATC
901 TTGATTGCCA TCGGTTTGTT TTTAATTTAC CAAAACGGGC TGACCCTGCT
951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGGCTC GGACTGCTGC
1001 CTATGCACAT CATCATGTTT GTCATCGCAA TCGTACTTCT GCGCGTCCGC
1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
35 1101 GAAAGgcgGA AAATGA

```

This corresponds to the amino acid sequence [<SEQ ID 506; ORF101ng-1>] (SEQ ID NO: 506; ORF101ng-1):

```

1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
40 51  LVGFWVIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101 PVMQFAVPFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLKD
201 NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
251 STAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
45 301 LIAIGLFLIY QNGLTLLFEA VEDGKIHF L GLLPMHIIMF VIAIVLLRVR
351 SMPSQPFWQA VGKSLTLKGG K*

```

ORF101ng-1 (SEQ ID NO: 506) and ORF101-1 (SEQ ID NO: 500) show 97.6% identity in 371 aa overlap:

-385-

		10	20	30	40	50	60
	orf101-1.pep	MIYQRNLIKELSFTAVGIFVVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT					
	orf101ng-1	MIYQRNLIKELSFTAVGIFVVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT					
5		10	20	30	40	50	60
	orf101-1.pep	70	80	90	100	110	120
	orf101ng-1	70	80	90	100	110	120
10		70	80	90	100	110	120
	orf101-1.pep	PLLLVLTAFTSTLTVLTRYWRDSEMSVWLSGLALKQWIRPVMQFAVPFAVLVAVMQLWV					
	orf101ng-1	PLLLVLTAFTSTLTVLTRYWRDSEMSVWLSGLALKQWIRPVMQFAVPFAVLVAVMQLWV					
		70	80	90	100	110	120
	orf101-1.pep	130	140	150	160	170	180
	orf101ng-1	130	140	150	160	170	180
15		130	140	150	160	170	180
	orf101-1.pep	IPWAE LRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ					
	orf101ng-1	IPWAE LRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ					
		130	140	150	160	170	180
	orf101-1.pep	190	200	210	220	230	240
	orf101ng-1	190	200	210	220	230	240
20		190	200	210	220	230	240
	orf101-1.pep	IDPVSHRRTIPTAQLIGSSNPQHQAELMWRI SLTVSVLLCLLAVPLSYFNPRSGHTYNI					
	orf101ng-1	IDPVSHRRTIPTAQLIGSSNPQHQAELMWRI SLTVSVLLCLLAVPLSYFNPRSGHTYNI					
25		250	260	270	280	290	300
	orf101-1.pep	LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFAVALILLRVRSMPSQPFWQA					
	orf101ng-1	LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFAVALILLRVRSMPSQPFWQA					
		250	260	270	280	290	300
	orf101-1.pep	310	320	330	340	350	360
	orf101ng-1	310	320	330	340	350	360
30		310	320	330	340	350	360
	orf101-1.pep	370	VGKSLTLKGGKX				
	orf101ng-1	370	VGKSLTLKGGKX				
35		370					

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 507] (SEQ ID NO: 507):

1 ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
 51 GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
 101 TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
 151 ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
 5 201 TCGTGTCGTC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
 251 CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
 301 AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
 351 TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
 10 401 ATTCGTAA

This corresponds to the amino acid sequence [<SEQ ID 508; ORF113>] (SEQ ID NO: 508; ORF113):

1 ..GGGFINASCA TLTTAKPQYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
 51 ILSYHISKIDA PVWGQDVRVV AGQNDVAATG DAHSPILNNA AANTSNTAN
 15 101 NGTHIPLFAI DTGKLGGXVC QQNHLDQYGR ASRHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein (SEQ ID NO: 1143) of *N.meningitidis* (accession AF030941)

20 ORF (SEQ ID NO: 508) and pspA (SEQ ID NO: 1143) show 44% aa identity in 179aa overlap:

orf113 GGGFINASCATLTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHISKIDA 60
 GGG INA+ TLT+ P G+L+ F + G VVI G GLD D DYTRILS ++I+A
 pspa GGGLINAASVTLTSGVPVLNNGNLTGPDVSSGKVIVGGKGLDTSADYTRILSRAAEINA 256
 orf113 PVWGQDVRVVAGQNDVAATGDAHSPILXXXXXXXXXXXXXXXXGTHIPLFAIDTGLKGGMYA 120
 25 VWG+DV+VV+G+N + G + P AIDT LGMYA
 pspa GVWGKDVKVVS GKNKLD FDG-----SLAKTASAPSSSDSVTPTVAIDTATLGGMYA 307
 orf113 NKITLISTVEQAGIRNQGWQFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
 +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A +++ A+ V N
 pspa DKITLISTDNGAVIRNKGRIFAATGGVTL SADGKLSNSGSIDAA----EITISAQTVDN 362

30 Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 (SEQ ID NO: 508) shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa overlap at the C-terminal part with a predicted ORF (ORF113ng) (SEQ ID NO: 510) from *N. gonorrhoeae*:

35 orf113 GGGFINASCATLTAKPQYQAGDLSAFKIR 30
 orf113ng SHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQAGDFSGFKIR 224
 orf113 QGNVVIAGHGLDARDTDYTRILSYHISKIDAPVWGQDVRVVAGQNDVAATGDAHSPILNNA 90
 orf113ng QGNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS 263

```

orf113                                IDTGKLGXVCQQNHLDDQYGRASRHS 135
                                     |||||:|||||
orf113ng    DFGSGFKIRQGNNAVIAGHGLDARDTDFTRILVCQQNHLDDQYGRTSRHS 263

```

The complete length ORF113ng nucleotide sequence [<SEQ ID 509>] (SEQ ID NO: 509) is
 5 predicted to encode a protein having amino acid sequence [<SEQ ID 510>] (SEQ ID NO: 510):

```

1  MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
51 SKAFCSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNIGP
101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
201 ASRATLTTGQ PQYQAGDFSG FKIRQGNNAVI AGHGLDARDT DFTRILVCQQ
251 NHLDDQYGRTS RHS*

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 511>] (SEQ ID NO: 511):

```

1  ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
51 CAACATTTCAT CTGGGTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
101 GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT
151 TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
201 ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
251 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGACAGC
301 CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGTTTATTA
351 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
401 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT
451 AATGGCGCGA CTGGCGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
501 AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
551 AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
601 CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
651 GTTGTCAAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
701 CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
751 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
801 ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
851 TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
901 CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
951 TATCACAGGC AAAGAAAAAG GTGTTT..

```

This corresponds to the amino acid sequence [<SEQ ID 512; ORF115>] (SEQ ID NO: 512; ORF115):

```

1  ..STGHSEQNYT LPREITRNI LGSFAYESHR KALSHHAPSQ GTELPQSNGI
51 SLPYTSNSFT PLPSSSLYII NPVNKGYLE TDPRFANYRQ WLGS DYMLDS
101 LKLDPNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
151 NGATAARSMN LSVGIALSAE QVAQLTSDIV WLWQKEVKLP DGGTQTVLVP

```

201 QVYVRVKNGD IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
 251 DNIGGRIHAQ KSAVTATQDI NNIGGMLSAE QTLLLNAGNN INSQSTTASS
 301 QNTQGSSTYL DRMAGIYITG KEKGV..

5 Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein (SEQ ID NO: 1143) of *N.meningitidis* (accession number AF030941)

ORF115 (SEQ ID NO: 512) and *pspA* protein (SEQ ID NO: 1143) show 50% aa identity in 325aa overlap:

10 Orf115: 1 STGHSEQNYTLPREITRNISLGSFAYESHKALSHHAPSQGTLPQSNGLSPYTSNSFT 60
 STG+S Y E++ +I +G AY+ + + P + NGI +T
 pspA: 778 STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVPVVAENGIHPTFT----- 831

Orf115: 61 PLPSSSLYIINPVNKGYLVEVDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEYEQR 120
 LP+SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEYEQ+
 15 pspA: 832 -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLGDGYEYEQK 890

Orf115: 121 LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
 L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
 pspA: 891 LVNEQIAKLTGYRRLDGYTNDEEQFKALMDNGITIAKELQTPGIALSAEQVARLTSDIV 950

Orf115: 181 WLQKEVKLPDGGTQTTLVLPQVYVRKNGDIDGKGALLSGSNTQINVSGSLKN-SGTIAG 239
 WL + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG
 20 pspA: 951 WLENETVTLPDGTTQTVLKPQVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAG 1009

Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLLLNAGXXXXXXXXXXXX 299
 R ALI+N + N+ G + + A DI N G + AE LLL A
 pspA: 1010 REALILNAQNIKNLQGDLOGKNIFAAAGSDITNTGS-IGAENALLLKASNNIESRSETRS 1068

25 Orf115: 300 XXXXXXXXXYLDRMAGIYITGKEKG 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorrhoeae*

ORF115 (SEQ ID NO: 512) shows 91.9% identity over a 334aa overlap with a predicted ORF

30 (ORF115ng) (SEQ ID NO: 514) from *N.gonorrhoeae*:

orf115.pep STGHSEQNYTLPREITRNISLGSFAYESHK 31
 ||| |||||:||||:||||||| |
 orf115ng NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHRQNYTLPEEITRDISLGSFAYESHK 71

orf115.pep ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVE 81
 |||:||||||| ||||| |||||:||||||| |
 35 orf115ng ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET 131

	orf115.pep	DPRFANYRQWLGS DYMLDSLKLD PNNLHKRLGDGYEQR LINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGDGYEQR LINEQIAELTGHRRLDGYQND	191
5	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVVLVQKEVKLPDGGTQTVLVPQ	201
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVVLVQKEVKLPDGGTQTVLMPQ	251
	orf115.pep	VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
	orf115ng	VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
10	orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
	orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
	orf115.pep	EKGV	325
15	orf115ng	EKGVLAAQAGKDINI IAGQISNQSDQGQTRLQAGRDINLDTVQTKGYQEIHFADADNHTIR	431

An ORF115ng nucleotide sequence [<SEQ ID 513>] (SEQ ID NO: 513) was predicted to encode a protein having amino acid sequence [<SEQ ID 514>] (SEQ ID NO: 514):

	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
20	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGS DYMLGS
	151	LKLD PNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLVQKEVKLP	DGGTQTVLMP
25	251	QVYVRVKNGG	IDGKGALLSG	SNTQINVSGS	LKNSGTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINI IAGQ	ISNQSDQGQT
	401	RLQAGRDINL	DTVQTKGYQE	IHFADADNHTI	RGSTNEVGSS	IQTGKDVTL
	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
30	501	GNKLVIDDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDTTIVAS	KHYEQTGSNV	SSPEGNLIS	TQSM DIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVTDLAQQ	AIAVAHKAAC	QFDKAKTTAL
	701	MPWRLPMQVG	RLFKQAKAPK	K*		

35 Further work revealed the following partial gonococcal DNA sequence [<SEQ ID 515>] (SEQ ID NO: 515):

	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAATGGTAA	GTTGCACAAC	TACTGGCGTG
	101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAACA	AAATTACT
40	151	TTGCCGAGG	AAATCACACG	CGACATTCA	CTGGGTCAT	TTGCCTATGA
	201	ATCGCATAGC	AAAGCATTAA	GCCGTCATGC	GCCCAGCCAA	GGCACTGAGT
	251	TGCCACAAAG	TAACCGGAT	AATATCCGTA	CTGCGAAAAG	CAACGGTATT
	301	TCGCTACCT	ATACGCCCAA	TTCTTTTACC	CCATTACCCG	GCAGCAGCTT
	351	ATACATTATC	AATCCTGCCA	ATAAAGGCTA	TCTTGTGAA	ACCGATCCAC
45	401	GCTTTGCCAA	CTACCGTCAA	TGGTTGGGTA	GTGACTATAT	GCTGGGCAGC
	451	CTCAAACTAG	ACCCAAACAA	TTACATAAA	CGTTTGGGTG	ATGGTTATTA
	501	CGAGCAACGT	TTAATCAATG	AACAAATCGC	AGAGCTGACA	GGGCATCGTC
	551	GTTTAGACGG	TTATCAAAAC	GACGAAGAAC	AATTTAAAGC	CTTAATGGAT
	601	AATGGCGCGA	CTGCGGCACG	TTCGATGAAT	CTCAGCGTTG	GCATTGCATT
50	651	AAGTGCCGAG	CAAGCAGCGC	AACTGACCAG	CGATATTGTT	TGGTTGGTAC

-390-

5 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
 801 GTTGTGAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
 901 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
 1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
 1101 TATCACAGGC AAAGAAAAAG GTGTTTATAG AGCGCAGGCA GGCAAAGACA
 10 1151 TCAACATCAT TGCCCGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
 1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAA
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCTatTG
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
 15 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
 20 1651 CAAGCAGGCA ATCATGTTCTG CATTTGTACA ACCCAAATC AAAGCCAAAG
 1701 CGAAACCTAT CATCAAACCC AAAATCAGG ATTGATGAGT GCAGGTATCG
 1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
 1851 TGTTCGAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
 25 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
 2001 GGTGGCATTG AGTTCGCCCG TTACCGATTG GGCACAACAA GCGATTGCCG
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
 30 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence [SEQ ID 516; ORF115ng-1] (SEQ ID NO: 516; ORF115ng-1):

35 1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
 51 LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPSQNRD NIRTAKSNGI
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
 151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP
 251 QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
 40 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN INNQSTAKSS
 351 QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINIAGQ ISNQSDQGQT
 401 RLQAGRDINL DTVQGTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLL
 451 SGNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
 501 GNKLVIDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI
 45 551 QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
 701 MPWRLPMQVG RPIKQAKAHK T*

50 This gonococcal protein (ORF115ng-1) (SEQ ID NO: 516) shows 91.9% identity with ORF115 (SEQ ID NO: 512) over 334aa:

20 30 40 50 60 70
 orf115ng-1.p NEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDISLGSFAYESHK
 55 orf115 STGHSEQNYTLPREITRNLISLGSFAYESHK

				10	20	30	
		80	90	100	110	120	130
	orf115ng-1.p	ALSRHAPSQGT	ELPQSNRDNIR	TAKSNGISLPY	TNSFTPLPGSS	LYIINPANKGY	LVET
5	orf115	ALSHHAPSQGT	ELPQSN-----	GISLPYTSNS	FTPLPSSSLY	IINPVNKG	YLVET
		40		50	60	70	80
		140	150	160	170	180	190
	orf115ng-1.p	DPRFANYRQWL	GSDYMLGSLK	LDPNNLHKRL	GDGYEQRLNE	QIAELTGHRRL	DGYQND
10	orf115	DPRFANYRQWL	GSDYMLDSLK	LDPNNLHKRL	GDGYEQRLNE	QIAELTGHRRL	DGYQND
		90	100	110	120	130	140
		200	210	220	230	240	250
	orf115ng-1.p	EEQFKALMDNG	AARSMNLSVG	IALSAEQAAQL	TSDIVWL	VQKEVKLPDGG	TQTVLMPQ
15	orf115	EEQFKALMDNG	AARSMNLSVG	IALSAEQVAQL	TSDIVWL	VQKEVKLPDGG	TQTVLVPQ
		150	160	170	180	190	200
		260	270	280	290	300	310
	orf115ng-1.p	VYVRVKNGGID	GKGALLSGSNT	QINVSGSLKNS	GTIAGRNALI	INTDTLDNIGG	RIHAQK
20	orf115	VYVRVKNGDID	GKGALLSGSNT	QINVSGSLKNS	GTIAGRNALI	INTDTLDNIGG	RIHAQK
		210	220	230	240	250	260
		320	330	340	350	360	370
	orf115ng-1.p	SAVTATQDINN	IGGILSAEQTL	LLNAGNNINN	QSTAKSSQNA	QGSSTYLDRM	AGIYITGK
25	orf115	SAVTATQDINN	IGGMLSAEQTL	LLNAGNNINS	QSTASSQNTQ	GSSTYLDRM	AGIYITGK
		270	280	290	300	310	320
		380	390	400	410	420	430
	orf115ng-1.p	EKGVLAAQAGK	DINIIAGQISN	QSDQGQTRLQ	AGRDLNLDTV	QTKYQEIHFDA	DNHTIR
30	orf115	EKGVLAAQAGK	DINIIAGQISN	QSDQGQTRLQ	AGRDLNLDTV	QTKYQEIHFDA	DNHTIR

In addition, it shows homology with a secreted *N.meningitidis* protein (SEQ ID NO: 1143) in the database:

35	gi 2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273
	Score = 604 bits (1541), Expect = e-172
	Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)
40	Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDIS 60
	L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
	Sbjct: 739 LIVGTPESALDNDDELGTCTI-TDKGDLHRYHRHKKGRDSTGYSRSPYEPAPEVS-SIR 796
	Query: 61 LGSFAYESHKALSRHAPSQGT
	+G AY+ + AP Q +++P + + NGI +T LP SSL+ I
	Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840
45	Query: 121 NPANKGYLVETDPRFANYRQWL
	P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT
	Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSYMLAALQDPNHIHKRLGDGYEQKLVEQIAKLT 900

Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLP 240
 G+RRRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
 Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLP 960

5 Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
 Sbjct: 961 DGGTQTVLKPQVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTY 359
 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
 Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

10 Query: 360 LDRMAGIYITGKEKGVLAQAQKDINIIAGQISNQSDQGQTRLQAGRNDINLDTVQTKGYQ 419
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLNAGNNLNAKAAEVGSAKGT LAVYAKNDITI 479
 FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
 15 Sbjct: 1139 NTIFDSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198

Query: 480 SSGIHAGQVDDASKHTGRSGGKNLVIDTKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

20 Query: 540 SNVISDNGTRIQAGNHVRIGTTQTQSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAATRSRRAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318

Query: 599 QSNEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658
 ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378

25 Query: 659 QTYEQKGLTVAFSSPVT 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QVYEQKGVTV AISVPV 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
 30 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 517>] (SEQ ID
 NO: 517):

35 1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
 51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
 101 ACACGACCCA TGTTGATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
 151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
 201 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT
 40 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG
 351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
 401 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT


```

501      TGTTCAGGC  AAACACTACG  AACAAATCGG  CAGTACCGTT  TCCAGCCCGG
551      AAGGCAACAA  TACCATCTAT  GCCCAAAGCA  TAGACATTCA  AGCGGCACAC
601      AACAAATTAA  ACAGTAATAC  CACCCAAACC  TATGAACAAA  AAGG.CTAAC
651      GGTGCATTTC  AGTTCGCCCC  TTACCGATTT  GGCACAACAA  ...

```

This corresponds to the amino acid sequence [[SEQ ID 518; ORF117](#)] ([SEQ ID NO: 518; ORF117](#)):

```

1      ..SGNNLNAAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
51     GNKLVIITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTQI
101    QAGNHVRIGT TQTQSQSEY HQTQKSGLS AGIGFTIGS TNTQENQSQS
151    NEHTGSTVGS LKGDITIVAG KHYEQIGLSTV SSPEGNNTIY AQSIDIQAAH
201    NKLSNNTTOT YEOKLTLVAF SSPVTDLAOO ...

```

Computer analysis of this amino acid sequence gave the following results:

15 Homology with the pspA putative secreted protein (SEQ ID NO: 1143) of *N.meningitidis*
(accession number AF030941)

ORF117 (SEQ ID NO: 518) and pspA protein (SEQ ID NO: 1143) show 45% aa identity in 224aa overlap:

20 Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVIITDKAQSH 63
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++
 pspA: 1173 DIRIRAAEVGSEQGRKLKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQ 1232

Orf117: 64 HETAQSSTFEGKQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQTSQSETYHQT 123
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
 pspA: 1233 NGQAVSGTLDGKEIILVSGRDIITVTGSNI IADNHTILSAKNNIVLKAAETR SRSAEMNKK 1292

25 Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGSCLKGDDTTIVAGKHYEQIGSTVSS 182
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGSL G+T I AGKHY Q GST+SS
 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352

Orf117: 183 PEGNNTIYAQSIDIQAAHNKLNSNTTQTYEQKXLTVAFSSPVT D 226
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
 30 pspA: 1353 PQGDVGISSGKISIDAAQNRYSQESKQVYEQKGVTV AISVPVVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 (SEQ ID NO: 518) shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) (SEQ ID NO: 520) from *N.gonorrhoeae*:

```

35      orf117.pep                               SGNNLNAKAAEVSSANGTLAVSANNDINIS      30
      |||||
      orf117nq      IHFDADNHTIRGSTNEVGSSIQTKGDVTLSSGNNLNAKAAEVGSAKGT LAVYAKNDITIS      480

```

5	orf117.pep	AGINTTHVDDASKHTGRSGGNGKLVIITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
	orf117ng	: ::: :	540
	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTKSGLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	: : : : : : :	600
	orf117.pep	NEHTGSTVGSCLKGDDTTIVAGKHYESIGSTVSSPEGNNNTIYAQSIDIQAHNKLNSTTQT	210
	orf117ng	: : : : : : :	660
10	orf117.pep	YEQKXLTVAFSSPVTDLAQQ	230
	orf117ng	: : : : : : :	720

An ORF117ng nucleotide sequence [<SEQ ID 519>] (SEQ ID NO: 519) was predicted to encode a
15 protein having amino acid sequence [<SEQ ID 520>] (SEQ ID NO: 520):

20	1	..LLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGSYMLGS
	151	LKLDPNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLQKEVKLP	DGGTQTVLMP
25	251	QVYVRVKNG	IDKGALLSG	SNTQINVSGS	LKNSCTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGLLSAE	QTLSTLGN	INNOSTAKSS
	351	QNAQSSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIAGQ	ISNQSDQGQT
	401	RLQAGRDINL	DTVQTKGYQE	IHFDADNHTI	RGSTNEVGSS	IQTKGDVTLL
	451	SGNNLNAKAA	EVGSAGGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
30	501	GNKLVIDTKA	QSHHETAQSS	TFEGQVVLQ	AGNDANILGS	NVISDNGTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDITIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSMDIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVTDLAQQ	AIAVAHKAAP	QFDKAKTTAL
	701	MPWRLPMQVG	RLFKQAKAPK	K*		

Further work revealed the following gonococcal partial DNA sequence [<SEQ ID 521>] (SEQ ID NO: 521):

35	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAATGGTAA	GTTGCACAAC	TACTGGCGTG
	101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAACA	AAATTATACT
	151	TTGCCGGAGG	AAATCACACG	CGACATTTC	CTGGGTTTAC	TTGCCTATGA
	201	ATCGCATAGC	AAAGCATTA	GCCGTCATGC	GCCACGCCAA	GGCATGTAGT
40	251	TGCCACAAAG	TAACCGGGAT	AATATCCGTA	CTCGGAAAAG	CAACCGTATT
	301	TCGCTACCCT	ATACGCCCAA	TTCTTTTACC	CCATTACCCG	GCAGCAGCTT
	351	ATACATTATC	AATCCTGCCA	ATAAAGGCTA	TCTTGTTGAA	ACCGATCCAC
	401	GCTTTGCCAA	CTACCGTCAA	TGGTTGGGTA	GTGACTATAT	GCTGGGCAGC
	451	CTCAAAC TAG	ACCCAAACAA	TTTACATAAA	CGTTTGGGTG	ATGGTTATTA
45	501	CGAGCAACGT	TTAATCAATG	AACAAATCGC	AGAGCTGACA	GGGCATCGTC
	551	GTTTAGACGG	TTATCAAAAC	GACGAAGAAC	AATTTAAAGC	CTTAATGGAT
	601	AATGGCGCGA	CTGCGGCACG	TTCGATGAAT	CTCAGCGTTG	GCATTGCATT
	651	AAGTGCCGAG	CAAGCAGCGC	AACTGACCAG	CGATATTGTT	TGGTTGGTAC
	701	AAAAAGAGT	TAAACTTCCT	GATGGCGGCA	CACAAACCGT	ATTGATGCCA
50	751	CAGGTTTATG	TACGCGTTAA	AAATGGCGGC	ATAGACGGTA	AAGGTGCATT
	801	GTTGTCAAGC	AGCAATACAC	AAATCAATGT	TTCAGGCAGC	CTGAAAAACT
	851	CAGGCACGAT	TGCAGGGCGC	AATGCGCTTA	TTATCAATAC	CGATACGCTA

5
10
15
20
25

```

901  GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
951  ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA
1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGCAAACC
1201 CGGCTGCAGG CAGGACGCGA CATTAACTG GATACGGTAC AAACCGGCAA
1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
1301 CGAACGAAGT CCGCAGCAGC ATTCAAACAA AAGGCGATGT TACCctatTG
1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
1451 ATGCCGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
1501 GGTAAATAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG
1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
1951 AACC AATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
2001 GGTGGCATTC AGTTCGCCG TTACCGATT GGCACAACAA GCGATTGCCG
2051 TAGCACACAA AGCAGCAAAAC AAGTCGACA AAGCAAAAC GACCGCGTTA
2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
2151 GGCGCACAAA ACTTAG

```

This corresponds to the amino acid sequence [SEQ ID 522; ORF117ng-1] (SEQ ID NO: 522; ORF117ng-1):

30
35
40
45

```

1  LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
51  LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
251 QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLLNAGNN INNQSTAKSS
351 QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINI IAGQ ISNQSDQGQT
401 RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL
451 SGNNLNAAA EVGSAGGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
501 GNKLVIDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI
551 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
601 NEHTGSTVGS LKGD TTIVAS KHYEQTGSNV SSPEGNNLIS TQSM DIGAAQ
651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
701 MPWR LPMQVG RPIKQAKAHK T*

```

ORF117ng-1 (SEQ ID NO: 522) shows the same 90% identity over a 230aa overlap with ORF117 (SEQ ID NO: 518). In addition, it shows homology with a secreted *N.meningitidis* protein (SEQ ID NO: 1143) in the database:

50

```

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length =
2273
Score = 604 bits (1541), Expect = e-172
Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

```

Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGLHNYWRARRKGHDETGHRQNYTLPEEITRDIS 60
L+V T + L N++T G K + ++ G L H Y R +K G D T G+ Y E++ I
Sbjct: 739 LIVGTPESALDNDLGTGTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796

5 Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
+G AY+ + AP Q +++P + + NGI +T LP SSL+ I
Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

10 Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELT 180
P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT
Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSGYMLAALQQDPNHIHKRLGDGYEQKLVNEQIAKLT 900

15 Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLP 240
G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLP 960

20 Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
Sbjct: 961 DGTQTQVLKPKVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

25 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTY 359
+ N+ G + + A D I N G I AE LLL A NNI ++S +S+QN QGS
Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASN NIESRSETRSNQNEQGSVRN 1078

30 Query: 360 LDRMAGIYITGKEKGVLAQAQKDINIIAGQISNQSDQGQTRLQAGRNDINLDTVQTKYQ 419
+ R+AGIY+TG++ G + AG +I + A +++NQs+ GQT L AG DI DT + Q
Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

35 Query: 420 EIHFADNHTIRGSTNEVGSSIQTKGDVTLNAGNNLNAKAAEVGSAKGT LAVYAKNDITI 479
FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
Sbjct: 1139 NTIFDS DNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198

40 Query: 480 SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
+G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDI TVTG 1258

45 Query: 540 SNVISDNGTRIQAQGNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETR SRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318

50 Query: 599 QSNEHTGSTVGS LKGD TTIVASKHYEQTSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658
++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRY SQESK 1378

55 Query: 659 QTYEQKGLTVAFSSPVT D 676
Q YEQKG+TVA S PV +
Sbjct: 1379 QVYEQKGVTV AISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 Example 63

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 523>] (SEQ ID NO: 523):

```

1  ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
5  51  CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC
151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCGGC
201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGyCATGCGC AACCTGCAAG
251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
10 351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA
451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
501 CGTGCGCATC GACTTCATCT CCTAT...
```

15 This corresponds to the amino acid sequence [<SEQ ID 524; ORF119>] (SEQ ID NO: 524; ORF119):

```

1  MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR
51  DGKPSGG SVM MPKPQPAVKK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP
20 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK
151 PLITLKLKLSK VELSWFDVRI DFISY...
```

Further work revealed the complete nucleotide sequence [<SEQ ID 525>] (SEQ ID NO: 525):

```

1  ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
25 51  CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAAACCAG CCATGTCCGC
151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCGGC
201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGCCATGCGC AACCTGCAAG
251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
30 351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC
401 CTGCCGACGC GCCGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA
451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA
501 CGTGCGCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCTGGC
35 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG
651 CTATCAGGCA TTTATCGTGG GTATTCAAGC AGTCAGCCGC AACCGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTGCGA
751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
40 851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGCAAAAAA CCTTCGACGA
45 1101 TTTGTTTATG GATTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGCCTGTT CTCCTAA
```

50 This corresponds to the amino acid sequence [<SEQ ID 526; ORF119-1>] (SEQ ID NO: 526; ORF119-1):

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
 51 DGKPSGGSSVM MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP
 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK
 151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
 201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
 251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
 301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
 401 RTYVLARQSE MLKVGIEPGG KTLRLFS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 (SEQ ID NO: 524) shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) (SEQ ID NO: 528) from strain A of *N. meningitidis*:

15		10	20	30	40	50	60
	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGKPSGGSSVM					
	orf119a	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM					
		10	20	30	40	50	60
20		70	80	90	100	110	120
	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
	orf119a	MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
		70	80	90	100	110	120
25		130	140	150	160	170	
	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY					
	orf119a	TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE					
		130	140	150	160	170	180
30	orf119a	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence [<SEQ ID 527>] (SEQ ID NO: 527) is:

1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA
 35 51 CAATATGTAT CAGGAAAACC AATACGCAA AAAAGTGCGC GACCAGTTCG
 101 GGCAC TCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACC GCG
 201 GGTCAAAAAA ACGGCAAAAT CCCAAGACCC CGCCATGCGC AACCTGCAAG
 251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 40 301 TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGGCA TTATCGGCAA
 351 CTCCGCCAC ACCGTTCCCG AACCCCAAAC CGGACATTCC GCACCAAAC
 401 CTGCCGACGC GCCGGCAAAA CCTGTTCCCG TTCCGCAAAC GCCGGCAAAA
 451 CCGCTGATTA CGCTCAAAGA GCTGTCAAG GTCGAGCTGC CCTGGTTTGA
 501 CGTGCGCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAAGAAC
 45 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCGGC
 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG
 651 TCATCAGGCA TTTATCGTGG GTATTCAAGC AGTCAGCCGC AACGGACTTG
 701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTGCA

5
10
15
20
25
30
35
40
45
50

751 CACAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGGAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGCGCCTGTT CTCCTAA

This encodes a protein having amino acid sequence [<SEQ ID 528>] (SEQ ID NO: 528):

15
20
25
30
35
40
45
50

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWFDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVVLARQSE MLKVGIEPGG KTAALRLFS*

ORF119a (SEQ ID NO: 528) and ORF119-1 (SEQ ID NO: 526) show 98.6% identity in 428 aa overlap:

30
35
40
45
50

orf119a.pep	10	20	30	40	50	60
orf119-1	10	20	30	40	50	60
orf119a.pep	70	80	90	100	110	120
orf119-1	70	80	90	100	110	120
orf119a.pep	130	140	150	160	170	180
orf119-1	130	140	150	160	170	180
orf119a.pep	190	200	210	220	230	240
orf119-1	190	200	210	220	230	240
orf119a.pep	250	260	270	280	290	300
orf119-1	250	260	270	280	290	300

-400-

		310	320	330	340	350	360
	orf119a.pep	AVTGVGVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
5	orf119-1	AVTGVGVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
		310	320	330	340	350	360
	orf119a.pep	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
10	orf119-1	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
		370	380	390	400	410	420
	orf119a.pep	KTALRLFSX					
	orf119-1	KTALRLFSX					
		429					

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 (SEQ ID NO: 524) shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) (SEQ ID NO: 530) from *N.gonorrhoeae*:

	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKVRDQFGHSDKDALLNSXTSHVRDQKPSGGSSVM	60
20	orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKVRDQFGHSDKDALLNSKTSHVRDQKPSGGPVM	60
	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEODAVYIAKQKQAKASPFKTEIETALEESGIIIGNSAH	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRNLQEODAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH	120
25	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY	175
	orf119ng	TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE	180

The complete length ORF119ng nucleotide sequence [<SEQ ID 529>] (SEQ ID NO: 529) is:

30	1	ATGATTTACA	TCGTACTGTT	CCTCGCCGCC	GTCCTCGCCG	TTGTCGCCTA
	51	CAATATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTGCGC	GACCAGTTCG
	101	GACACTCCGA	CAAAGATGCC	CTGCTCAACA	GCAAAACCAG	CCATGTCCGC
	151	GACGGCAAAC	CGTCCGGCGG	GCCAGTCATG	ATGCCGAAAC	CCCAACCGGC
	201	GGTCAAAAAA	CCGGCCAAAC	CCCAAGACTC	CGCCATGCGC	AACCTGCAAG
	251	AACAGGATGC	CGTCTACATC	GCCAAGCAGA	AACAGGCAAA	AGCCTCCCCG
35	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGA	GAAATCGGCA	TTATCGGCAA
	351	CTCCGCCCAC	ACCGTTTCCG	AACCCCAAAC	CGGACATTCC	GCACCGAAAC
	401	CTGCCGACGC	GCCGGCAAAA	CCCGTTCCCG	TTCCGCAAAC	GCCGGCAAAA
	451	CCGCTGATTA	CGCTCAAAGA	GCTGTGCAAG	GTCGAGCTGC	CCTGGTTTGA
	501	CGTGCGCTtc	gACTTCATCT	CCTATATCGC	GCTGACCGAA	GCCAAAGAAC
40	551	TGCACGCACT	GCCGCGCCTT	tccAACCGCT	GCCGCTACCA	GATTGTGCGC
	601	TGCACCATGG	ACGACCATT	CCAGATTGCC	GAACCCATCC	CGGGCATCCG
	651	CTATCAGGCA	TTATCGTGG	GTATCCAGGC	AGTCAGCCGC	AACGGACTTG
	701	CCTCGCAGGA	AGAACTCTCC	GCATTCAACC	GCCAGGCGGA	CGCATTCGCA
	751	CAAAGCATGG	GCGGTCAAG	GCTGCACACC	GACCTTGCCG	CCTTTATCGA
45	801	AGTGGCTTCC	GCACTGGACG	CATTCTGCGC	GCGCGTCGAC	CAGACCATCG
	851	CCATCCATTT	GGTTTCGCCG	ACCAGCATCA	GCGGCGTAGA	ACTGCGTTCC
	901	GCCGTAACGG	GCGTGGGTTT	CGTTTGGGAA	GACGACGGCG	CGTTCCACTA
	951	TACCGACACG	TCGGGCTCGA	CCATGTTCTC	CATCTGCTCG	CTCAACAACG

5

1001	AGCCGTTTAC	CAATGCCCTT	TTGGACAACC	AGTCTTACAA	AGGCTTCAGT
1051	ATGCTGCTCG	ACATCCC GCA	CTCTCCGGCA	GGCGAAAAAA	CCTTCGACGA
1101	TTTGTTTATG	GATTTGGCGG	TACGCCTGTC	CGGTCAAGTTG	AACCTGAATC
1151	TGGTCAACGA	CAAAATGGAA	GAAGTTTCGA	CCCAATGGCT	CAAAGACGTA
1201	CGCACTTATG	TATTGGCGCG	TCAGTCCGAG	ATGATCAAAG	TCGGTATCGA
1251	ACCGGGCGGC	AAAACCGCCC	TGCGCCTGTT	TTCATAA	

This encodes a protein having amino acid sequence [<SEQ ID 530>] (SEQ ID NO: 530):

10	1	MIYIVLFLAA	VLAVVAYNMY	QENQYRKVKR	DQFGHSDKDA	LLNSKTSHVR
	51	DGKPSGGPVM	MPKPQPAVKK	PAKPQDSAMR	NLQEQDAVYI	AKQKQAKASP
	101	FKTEIETALE	EIGIIGNSAH	TVSEPQTGHS	APKPADAPAK	PVPVPQTPAK
	151	PLITLKELSK	VELPWFDVRF	DFISYIALTE	AKELHALPRL	SNRCRYQIVG
	201	CTMDDHFQIA	EPPIGIRYQA	FIVGIQAVSR	NGLASQEEELS	AFNRQADAFa
15	251	QSMGGQTLHT	DLAAFIEVAS	ALDAFCARVD	QTIAIHLVSP	TSISGVELRS
	301	AVTGVGVFLE	DDGAFHYTDT	SGSTMFSICS	LNNPFTNAL	LDNQSYKGS
	351	MLLDIPHSFA	GKFTFDDLFM	DLAVRLSGQL	NLNLVNDKME	EVSTQWLKDV
	401	RTYVLAROSE	MLKVGIEPGG	KTALRLFS*		

ORF119ng (SEQ ID NO: 530) and ORF119-1 (SEQ ID NO: 526) show 98.4% identity over 428 aa overlap:

		10	20	30	40	50	60
	orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVVDGKPSGGPVM					
	orf119-1	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVVDGKPSGGSV					
25		10	20	30	40	50	60
	orf119ng	MPKPQPAVKKPAKQDSAMRNLEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH					
	orf119-1	MPKPQPAVKKTAKPQDPAMRNLEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
30		70	80	90	100	110	120
	orf119ng	TVSEPQTGHSAKPADAPAKPVVPQTPAKPLITLKELSKVELPWFVDFRDFISYIALTE					
	orf119-1	TVSEPQTGHSAKPADAPAKPAPVPQTPAKPLITLKELSKVELPWFVDFRDFISYIALTE					
35		130	140	150	160	170	180
	orf119ng	AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVG IQAVSRNGLASQEELS					
	orf119-1	AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVG IQAVSRNGLASQEELS					
40		190	200	210	220	230	240
	orf119ng	AFNRQADAFQAQSMGGQTLHTDLAAFEIVASALDAFCARVDQTTIAIHLVSPTSISGVELRS					
	orf119-1	AFNRQVDAFAQSMGGQTLHTDLAAFEIVASALDAFCARVDQTTIAIHLVSPTSISGVELRS					
45		250	260	270	280	290	300
	orf119ng	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
	orf119-1	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
50		310	320	330	340	350	360
	orf119ng	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
	orf119-1	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					

```

          370      380      390      400      410      420
orf119ng  GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG
          |||||
orf119-1  GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG
          370      380      390      400      410      420

          429
orf119ng  KTALRLFSX
          |||||
orf119-1  KTALRLFSX

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 64

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 531] (SEQ ID NO: 531)

```

1  ..GCGCGGCACG GCACGGAAGA TTTCTTCATG AACAAACAGCG ACAC.ATCAG
51  GCAGATAGTC GAAAGCACCA CCGGTACGAT GAAGCTGCTG ATTTCTCTCCA
101 TCGCCCTGAT TTCATTGGTA GTCGGCGGCA TCGGCGTGAT GAACATCATG
151 CTGGTGTCCG TTACCGAGCG CACCAAAGAA ATCGGCATAC GGATGGCAAT
201 CGGCGCGCGG CGCGGCAATA TTTyGCAGCA GTTTTGTGATT GAGGCGGTGT
251 TAATCTGCGT CATCGGCGGT TTGGTCGGCG TGGGTTTGTC CGCCGCCGTC
301 AGCCTCGTGT TCAATCATTT TGTAACCGAC TTCCCGATGG ACATTTCCGC
351 CATGTCCGTC ATCGGCGCGG TCGCCTGTTC GACCGGAATC GGCATCGCGT
401 TCGGCTTTAT GCCTGCCAAT AAAGCAGCCA AACTCAATCC GATAGACGCA
451 TTGGCACAGG ATTGA

```

This corresponds to the amino acid sequence [SEQ ID 532; ORF134] (SEQ ID NO: 532; ORF134):

```

1  ..ARHGTEDFFM NNSDXIRQIV ESTTGTMKLL ISSIALISLV VGGIGVMNIM
51  LVSVTERTKE IGIRMAIGAR RGNIXQQLI EAVLICVIGG LVGVGLSAAV
101 SLVFNHFVTD FPMDISAMSV IGAVACSTGI GIAFGFMPAN KAAKLNPIDA
151 LAQD*

```

Further work revealed the complete nucleotide sequence [SEQ ID 533] (SEQ ID NO: 533):

```

1  ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
51  GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTCGCTC GTCGCATTGG
101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
151 AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGCGGACA GGCAGCGCGG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
351 TTTCGACGTG CGCGGACTGA AGCTGGAAC GGGGCGGCTG TTTGACGAAA
401 ACGATGTGAA AGAAGACGCG CAGGTCGTCTG TCATCGACCA AAATGTCAAA
451 GACAACTCT TTGCGGACTC GGATCCGTTG GGTAAACCA TTTGTTCAG

```

5 501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAACGCTT
 551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCTATAC GACGGTGATG
 601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
 651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
 701 AAGCGCGGCA CGGCACGGAA GATTCTTCA TGAACAACAG CGACAGCATC
 751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCTC
 801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
 851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
 10 901 ATCGGCGCGC GGCGCGGCAA TATTTGTCAG CAGTTTTGA TTGAGGCGGT
 951 GTTAATCTGC GTCATCGGCG GTTGGTTCGG CGTGGGTTTG TCCGCCGCCG
 1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
 1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
 1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAACTCAAT CCGATAGACG
 1151 CATTGGCACA GGATTGA

This corresponds to the amino acid sequence [<SEQ ID 534; ORF134-1>] (SEQ ID NO: 534; ORF134-1):

20 1 MSVQAVLAHK MRSLLTMLGI IIGIASVVSV VALGNGSQKK ILEDISSIGT
 51 NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTTLT
 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
 151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
 201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNNSDSI
 251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
 301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
 25 351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 (SEQ ID NO: 1144) of *E.coli* (accession number AE000189)

30 ORF134 (SEQ ID NO: 532) and o648 protein (SEQ ID NO: 1144) show 45% aa identity in 153aa overlap:

35 Orf134: 2 RHGTEDFFMNNSDXIRQIVESTTGTMKXXXXXXXXXXXXVVGGIGVMNIMLVSVTERTKEI 61
 RHG +DFF N D + + VE TT T++ VVGGIGVMNIMLVSVTERT+EI
 o648: 496 RHGKKDFFTWNMDGVLKTVEKTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREI 555
 Orf134: 62 GIRMAIGARRGNIXQQFLIEAXXXXXXXXXXXXXXXXXXXXXFNHFVTDFPMDISAMSVI 121
 GIRMA+GAR ++ QQFLIEA F+ + + S ++++
 o648: 556 GIRMAVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALL 615
 Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154
 A CST GI FG++PA AA+L+P+DALA++
 40 o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 (SEQ ID NO: 532) shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) (SEQ ID NO: 536) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF134a nucleotide sequence [<SEQ ID 535>] (SEQ ID NO: 535) is:

25	1	ATGTCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTTCG	TTCTGACGAT
	51	GCTCGGCATC	ATCATCGGTA	TCGCTTCGGT	TGTCCTCCGT	GTCGCATTGG
	101	GCAACGGTTC	GCAGAAAAAA	ATCCTTGAAG	ACATCAGTTC	GATAGGGACG
	151	AACACCATCA	GCATCTTCCC	AGGGCGCGCG	TTCCGGCGACA	GGCGCAGCGG
	201	CAGGATTAA	ACCTTGACCA	TAGACGACGC	AAAAATCATC	GCCAAACAAA
30	251	GCTACGTTGC	TTCCGCCACG	CCCATGACTT	CAGCGGCGCG	CACCGTGACT
	301	TACCGCAATA	CCGACCTGAC	CGCTTCTTTG	TACGGTGTGG	GCGAAACAATA
	351	TTTCGACGTG	CGCGGGCTGA	AGCTGGA AAC	GGGGCGGCTG	TTTGACGAAA
	401	ACGATGTGAA	AGAAGACGCG	CAGGTCGTCT	TCATCGACCA	AAATGTCAAAA
	451	GACAAACTCT	TTGCGGACTC	GGATCCGTTG	GGTAAAACCA	TTTTTGTTAG
35	501	GAAACGCCCC	TTGACCGTGA	TCGGCGTGAT	GAAAACAGAC	GAAAACGCTG
	551	TCGGCAATTC	CGACGTGCTG	ATGCTTTGGT	CGCCCTATAC	GACCGTGATG
	601	CACCAAAATCA	CAGGCGAGAG	CCACACCAAC	TCCATCACCG	TCAAAATCAA
	651	AGACAATGCC	AATACCCAGG	TTGCCGAAAA	AGGGCTGACC	GATCTGCTCA
	701	AAGCGCGGCA	CGGCACGGAA	GATTTCTTCA	TGAACAAACG	CGACAGCATC
40	751	ATGCAGATAG	TCGAAAGCAC	CACCGGTACG	ATGAAGCTGC	TGATTTCTCT
	801	CATCGCCCTG	ATTTTATTGG	TAGTCGGCGG	CATCGGCGTG	ATGAACATCA
	851	TGCTGGTGTC	GCTTACCGAG	CGCACCAAAG	AAATCGGCAT	ACGGATGGCA
	901	ATCGGCGCGC	GGCGCGGCAA	TATTTTGCAG	CAGTTTTTTGA	TTGAGGCGGT
	951	GTTAATCTGC	GTCATCGGCG	GTTTGGTCGG	CGTGGGTTTG	TCCGCCGCCG
45	1001	TCAGCCTCGT	GTTCAATCAT	TTTGTAAACG	ACTTCCCGAT	GGACATTTCC
	1051	GCCATGTCCG	TCATCGGCGC	GGTCGCCTGT	TCGACCGGAA	TCGGCATCGC
	1101	GTTCCGGTTT	ATGCTGCGCA	ATAAAGCAGC	CAAACCTCAAT	CCGATAGATG
	1151	CATTGGCGCA	GGATTGA			

This encodes a protein having amino acid sequence [<SEQ ID 536>] (SEQ ID NO: 536):

50 1 MSVOAVLAHK MRSLLTMLGI IIGIASVVSV VALGNCSOKK ILEDISSIGT

51 NTISIFPGRG FGDRRSGRK TLTIDDAKII AKQSYVASAT PMTSSGGTTLT
 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
 151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
 201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
 251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
 301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
 351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

10 ORF134a (SEQ ID NO: 536) and ORF134-1 (SEQ ID NO: 534) show 100.0% identity in 388 aa overlap:

15 orf134a.pep MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALGNGSQKKILEDISSIGTNTISIFPGRG
 orf134-1 MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALGNGSQKKILEDISSIGTNTISIFPGRG
 orf134a.pep FGDRRSGRKTLTIDDAKIIAKQSYVASATPMTSSGGTLYRNTDLTASLYGVGEQYFDV
 orf134-1 FGDRRSGRKTLTIDDAKIIAKQSYVASATPMTSSGGTLYRNTDLTASLYGVGEQYFDV
 orf134a.pep RGLKLETGRLENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
 orf134-1 RGLKLETGRLENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
 20 orf134a.pep ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE
 orf134-1 ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE
 orf134a.pep DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
 25 orf134-1 DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
 orf134a.pep IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
 orf134-1 IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
 orf134a.pep STGIGIAFGFMPANKAAKLNPIDALAQDX
 30 orf134-1 STGIGIAFGFMPANKAAKLNPIDALAQDX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF134 (SEQ ID NO: 532) shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) (SEQ ID NO: 538) from *N. gonorrhoeae*:

35 orf134.pep ARHGTEFFMNSDXIRQIVESTTGTMKLL 30
 orf134ng GESHTNSITVKIKDNANTRVAEKGLAELLKARHGTEFFMNSDSIRQMVESTTGTMKLL 264
 orf134.pep ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG 90
 40 orf134ng ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG 324

```

orf134.pep  LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA  150
             |||||
orf134ng    LVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVACSTGIGIAFGFMPANKAAKLNPIDA  384

orf134.pep  LAQD  154
             ||||
orf134ng    LAQD  388

```

The complete length ORF134ng nucleotide sequence [<SEQ ID 537>] (SEQ ID NO: 537) is:

```

10      1  ATGTCGGTGC  AAGCAGTATT  GGCGCACAAA  ATGCGTTCGC  TTCTGACCAT
      51  GCTCGGCATC  ATCATCGGTA  TCGCTTCGGT  TGTCTCCGTC  GTCGCGCTGG
     101  GCAACGGTTC  GCAGAAAAAA  ATCCTCGAAG  ACATCAGTTC  GATGGGGACG
     151  AACACCATCA  GCATCTTCCC  CGGGCGCGGC  TTCGGCGACA  GGCGCAGCGG
     201  CAAAATCAAA  ACCCTGACCA  TAGACGACGC  AAAAATCATC  GCCAAACAAA
     251  GCTACGTTGC  CTCCGCCACG  CCCATGACTT  CGAGCGGCGG  CACGCTGACC
    301  TACCGCAATA  CCGACCTGAC  CGCTTCTTTG  TACGGTGTGG  GCGAACAATA
    351  TTTCGACGTG  CGCGGGCTGA  AGCTGGA AAC  GGGGCGGCTG  TTTGATGAGA
    401  ACGATGTGAA  AGAAGACGCG  CAAGTCGTCG  TCATCGACCA  AAATGTCAAA
    451  GACAAACTCT  TTGCGGACTC  GGATCCGTTG  GGTAAAACCA  TTTTGTTTCA
    501  GAAACGCCCC  TTGACCGTCA  TCGGCGTGAT  GAAAAAAGAC  GAAAACGCTT
    551  TCGCAATTC  CGACGTGCTG  ATGCTTTGGT  CGCCCTATAC  GACGGTGATG
    601  CACCAATCA  CAGGCGAGAG  CCACACCAAC  TCCATCACCG  TCAAAATCAA
    651  AGACAATGCC  AATACCCGGG  TTGCCGAAAA  AGGGCTGGCC  GAGCTGCTCA
    701  AAGCACGGCA  CGGCACGGAA  GACTTCTTTA  TGAACAACAG  CGACAGCATC
    751  AGGCAGATGG  TCGAAAGCAC  CACCGGTACG  ATGAAGCTGC  TGATTTCTCT
    801  CATCGCCCTG  ATTCATTGG  TAGTCGGCGG  CATCGGTGTG  ATGAACATTA
    851  TGCTGGTGTC  CGTTACCGAG  CGCACCAAAG  AAATCGGCAT  ACGGATGGCA
    901  ATCGGCGCGC  GCGCGGCAA  TATTTTGCAG  CAGTTTTTGA  TTGAGGCGGT
    951  GTTAATCTGC  ATCATCGGAG  GCTTGGTCGG  CGTAGGTTTG  TCCGCCGCGG
   1001  TCAGCCTCGT  GTTCAATCAT  TTTGTAACCG  ATTTCCCGAT  GGACATTTCG
   1051  GCGGCATCCG  TTATCGGGGC  GGTGCGCTGT  TCGACCGGAA  TCGGCATCGC
   1101  GTTCGGCTTT  ATGCCTGCCA  ATAAGGCAGC  CAAACTCAAT  CCGATAGATG
   1151  CATTGGCGCA  GGATTGA

```

This encodes a protein having amino acid sequence [<SEQ ID 538>] (SEQ ID NO: 538):

```

35      1  MSVQAVLAHK  MRSLLTMLGI  IIGIASVSV  VALNGSQKK  ILEDISSMGT
      51  NTISIFPGRG  FGDRRSGKIK  TLTIDDAKII  AKQSYVASAT  PMTSSGGTLT
     101  YRNTDLTASL  YGVGEQYFDV  RGLKLETGRL  FDENDVKEDA  QVVVIDQNVK
     151  DKLFADSDPL  GKTILFRKRP  LTVIGVMKKD  ENAFGNSDVL  MLWSPYTTVM
     201  HQITGESHTN  SITVKIKDNA  NTRVAEKGLA  ELLKARHGTE  DFFMNSDSI
     251  RQMVSTTGT  MKLLISSIAL  ISLVVGGIGV  MNIMLVSVTE  RTKEIGIRMA
     301  IGARRGNILQ  QFLIEAVLIC  IIGGLVGVGL  SAAVSLVFNH  FVTDFPMDIS
     351  AASVIGAVAC  STGIGIAFGF  MPANKAAKLN  PIDALAQD*

```

ORF134ng (SEQ ID NO: 538) and ORF134-1 (SEQ ID NO: 534) show 97.9% identity in 388 aa

45 overlap:

```

50      orf134ng    MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGSQKKILEDISSMGTNTISIFPGRG
             |||||
      orf134-1    MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGSQKKILEDISSIGTNTISIFPGRG

      orf134ng    FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
             |||||:|||||
      orf134-1    FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

```

```

5  orf134ng      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

   orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKKIDNANTRVAEKGLAELLKARHGTE
   orf134-1      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKKIDNANTQVAEKGLTDLKARHGTE

   orf134ng      DFFMNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134-1      DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

10  orf134ng      IGARRGNILQQFLIEAVLICIIIGLVGVGLSAAVSLVFNHFVTDFPMDISAASVIGAVAC
   orf134-1      IGARRGNILQQFLIEAVLICVIGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC

   orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
15  orf134-1      STGIGIAFGFMPANKAAKLNPIDALAQDX

```

ORF134ng (SEQ ID NO: 538) also shows homology to an *E.coli* ABC transporter (SEQ ID NO: 1145):

```

20  sp|P75831|YBJZ_ECOLI_HYPOTHETICAL_ABC_TRANSPORTER_ATP-BINDING_PROTEIN_YBJZ_gi5
   (AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length = 648
   Score = 297 bits (753), Expect = 6e-80
   Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

   Query: 1  MSVQAVLAHKMRSLTMLXXXXXXXXXXXXXXXXXLGNGSQKKILEDISSMGNTNTISIFPGRG 60
   M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
25  Sbjct: 260 MAWRALAANKMRTLTLMLGIIIGIASVVSIVVVGDAKQMVLAIRSIGTNTIDVYPGKD 319

   Query: 61  FGDRRSGBKIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
   FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
   Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

   Query: 121 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179
   G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
30  Sbjct: 380 YGMTFSEGNTFNQEQLNGRAQVVVLDNTRRQLFPHKADVGEVILVGNMPARVIGVAEE 439

   Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKKIDNANTRVAEKGLAELLKARHGT 239
   ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
   Sbjct: 440 KQSMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRVKEGFDSEAEQQLTRLLSLRHGK 499

   Query: 240 EDFFMNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEIGIRM 299
   +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EIGIRM
35  Sbjct: 500 KDFFTWNMDGVLTVEKTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREIGIRM 559

   Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFPMDISAASVIGAVA 359
   A+GAR ++LQQFLIE F+ + + S +++ A
40  Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619

   Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
   CST GI FG++PA AA+L+P+DALA++
   Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 539>] (SEQ ID NO: 539):

```

1  ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCCTTT
51  GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCCCTTCCT GATTTTGAAG GAACGGATTT CCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGC CGATGT
251 CCGGCTGGGC GTATTTGAAG GTGCGGAAC TGTCTTTGGC GGGCGAACCC
301 GGCTGGCGCG TCGTGTTTTA CCTTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
15 401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA
501 TATGACCGTC GTTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA
20

```

This corresponds to the amino acid sequence [<SEQ ID 540; ORF135>] (SEQ ID NO: 540; ORF135):

```

1  ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVFSFLILK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
25 101 GWRVVFYLSV TGVMSSVWA TLTGWHTLSF PSAVYLSLIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELFWQEI LGMCIISAV
201 F*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 541>] (SEQ ID NO: 541):

```

30 1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GGCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
35 251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACTGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTGG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGTTT
401 TTGCCGCGT GGTATTGCTG CTTAATCCCT CGTCCGCAG CGGTCAGGAA
451 ACGGCGGCAC TCGCCGGGCT GCGGGCGGCG GCGATGTCCG GCTGGGCGTA
40 501 TTTGAAAGTG CGGGAACGT CTTTGCGCGG CGAACCCGCG TGCGCGCTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTGTCGGT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGCGGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
701 AAGTCGGCGA CAAATTCACG GTTGCTCGC TTTCTATAT GACCGTCGTT
45 751 TTTTCCGCTC TGTCTGCCG ATTTTCTCTG GCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA

```


851 TCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

This corresponds to the amino acid sequence [<SEQ ID 542; ORF135-1>] (SEQ ID NO: 542;

5 ORF135-1):

10

1	<u>MDTAKKDILG</u>	<u>SGWMLVAAAC</u>	<u>FTIMNVLIKE</u>	<u>ASAKFALGSG</u>	<u>ELVFWRMLFS</u>
51	<u>TVALGAAAVL</u>	<u>RRDXFRTPHW</u>	<u>KNHLNRSMVG</u>	<u>TGAMLLLFYA</u>	<u>VTHLPLATGV</u>
101	<u>TLSYTSSIFL</u>	<u>AVFSFLILKE</u>	<u>RISVYTQAVL</u>	<u>LLGFAGVVLL</u>	<u>LNPFSRSGQE</u>
151	<u>TAALAGLAGG</u>	<u>AMSGWAYLKV</u>	<u>RELSLAGEPG</u>	<u>WRVVFYLSVT</u>	<u>GVAMSSVWAT</u>
201	<u>LTGWHTLSP</u>	<u>SAVYLSICIGV</u>	<u>SALIAQLSMT</u>	<u>RAYKVGDKFT</u>	<u>VASLSYMTVV</u>
251	<u>FSALSAAFFL</u>	<u>GEELFWQEIL</u>	<u>GMCIILSGI</u>	<u>LSSI RPTAFK</u>	<u>QRLQSLFRQR</u>
301	*				

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 (SEQ ID NO: 540) shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) (SEQ ID NO: 544) from strain A of *N. meningitidis*:

```

20      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
                        |||||
orf135a      STVALGAAVLRRTDFRTPHWKNHLNRSMTGAMLLLFYAVTHLPLATGVTLSTSSIF
              50      60      70      80      90      100

25      orf135.pep      40      50      60      70      80      90
                        LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
                        |||||
orf135a      LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
              110     120     130     140     150     160

30      orf135.pep      100     110     120     130     140     150
                        VRELSSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSICIGVSALIAQLSM
                        |||||
orf135a      VRELSSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSICIGVSALIAQLSM
              170     180     190     200     210     220

35      orf135.pep      160     170     180     190     200
                        TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVFX
                        |||||
orf135a      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLAEEELFWQEILGMCIIILSGILSSIRPTAF
              230     240     250     260     270     280

40      orf135a      KQRLQSLFRQRX
              290     300

```

The complete length ORF135a nucleotide sequence [<SEQ ID 543>] (SEQ ID NO: 543) is:

1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51 GGC GGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA

-410-

5
10
15

```

101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC
201 GCCCCATTGG AAAAACCAC TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGGTT
401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
451 ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG
601 CTGACCGGCT GGCACACCTT GTCCTTTCCA TCGGCAGTTT ATCTGTCTGT
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
701 AAGTCGGCGA CAAATTCACG GTTGCTCGC TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCG ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

This encodes a protein having amino acid sequence [<SEQ ID 544>] (SEQ ID NO: 544):

20
25
30

```

1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
51 TVALGAAAVL RRDTRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTTV
251 FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

ORF135a (SEQ ID NO: 544) and ORF135-1 (SEQ ID NO: 542) show 99.3% identity in 300 aa overlap:

30
35
40

```

orf135a.pep MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

orf135a.pep RRDTRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
orf135-1 RRDTRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE

orf135a.pep RISVYTQAVLLLGFGAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
orf135-1 RISVYTQAVLLLGFGAGVVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG

orf135a.pep WRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMTRAYKVGDKFT

orf135a.pep VASLSYMTTVFSALSAAFFLAELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
orf135-1 VASLSYMTTVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

```

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 (SEQ ID NO: 540) shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) (SEQ ID NO: 546) from *N.gonorrhoeae*:

```

5      orf135.pep      GTGAMLLLFFYAVTXLPLATGVTLSYTSSIF      30
      orf135ng      STVTLGAAAVLRRDTRTPHWKNHLNRSMTGAMLLLFFYAVTHLPLTTGVTLSYTSSIF      335
      orf135.pep      LAVFSFLILKERISVYTQAVLLLGFAGVLLLLNPSFRSGQETAALAGLAGGAMSGWAYLK      90
      orf135ng      LAVFSFLILKERISVYTQAVLLLGFAGVLLLLNPSFRSGQEPALAGLAGGAMSGWAYLK      395
10     orf135.pep      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSM      150
      orf135ng      VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSM      455
      orf135.pep      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF      201
      orf135ng      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF      506
15

```

An ORF135ng nucleotide sequence [<SEQ ID 545>] (SEQ ID NO: 545) was predicted to encode a protein having amino acid sequence [<SEQ ID 546>] (SEQ ID NO: 546):

```

20     1  MPSEKAFRRH  LRTASFQGLH  LHHFHQKVGK  CGIIGFGIHI  FPTLLPAAQG
      51  ILDIQLGLFR  IDFAALAVYR  RTQVDFIHTV  IDGIASDQAF  SEVVQILRRL
      101 NLGHFTDTHL  IAQARRFIAD  FGNIRPMRRG  EAKTFCRCFR  FDGIDGIHGD
      151 FRQCGHINRL  APGKDCRNGK  RDKVFFHTRH  YNQVCLEKTN  CSARKIKFRH
      201 QKQAKTHSTS  LAARFTIRPS  LSQRPFMDTA  KKDILGSGWM  LVAAACFTVM
      251 NVLIKEASAK  FALGSGELVF  WRMLFSTVTL  GAAAVLRRDT  FRTPHWKNHL
      301 NRSMTGAM    LLLFYAVTHL  PLTTGVTLSY  TSSIFLAVFS  FLILKERISV
25     351 YTQAVLLLGF  AGVLLLLNPS  FRSGQEPAL  AGLAGGAMSG  WAYLKVRELS
      401 LAGEPGWRV  FYLSATGVAM  SSVWATLTGW  HTLSFPSAVY  LSGIGVSALI
      451 AQLSMTRAYK  VGDKFTVASL  SYMTVVFSAL  SAAFFLGEEL  FWQEILGMCI
      501 IISAAF*

```

30 Further work revealed the following gonococcal sequence [<SEQ ID 547>] (SEQ ID NO: 547):

```

35     1  ATGGATACCG  CAAAAAAGA  CATTTTAGGA  TCGGGCTGGA  TGCTGGTGGC
      51  GCGGCGCTGC  TTCACCGTTA  TGAACGTATT  GATTAAAGAG  GCATCGGCAA
      101 AATTTGCCCT  CGGCAGCGGC  GAATTGGTCT  TTTGGCGCAT  GCTGTTTTCA
      151 ACCGTTACGC  TCGGTGCTGC  CGCCGTATTG  CGGCGCGACA  CCTCCGCAC
      201 GCCCCATTGG  AAAAACCCT  TAAACCGCAG  TATGGTCGGG  ACGGGGGCGA
      251 TGCTGCTGCT  GTTTTACGCG  GTAACGCATC  TGCCTTTGAC  AACCGGCGTT
      301 ACCCTGAGTT  ACACCTCGTC  GATTTTTttg  GCGGTATTTT  CCTTCCTGAT
      351 TTTGAAAGAA  CGGATTTCCG  TTTACACGCA  GCGGTGCTG  CTCCTTGTT
      401 TTGCCGCGCT  GGTATTGCTG  CTTAATCCCT  CGTTCCGCAG  CGGTCAGGAA
40     451 CCGGCGGCAC  TCGCCGGGCT  GCGGGCGGC  GCGATGTCCG  GCTGGGCGTA
      501 TTTGAAAGTG  CGCGAAGTGT  CTTTGGCGGG  CGAACCCGGC  TGGCGCGTCG
      551 TGTTTTACCT  TTCCGCAACC  GGCGTGCGCA  TGTCGTCggt  ttggcgacg
      601 Ctgaccggt  ggCACAcccT  GTCCTTTcca  tcggcagttt  ATCtgtCGGG
      651 CATCGGCGTG  tccgcgCtgA  TTGCCCAaCT  GtcgatgAcg  cGCGcctaca
45     701 aaGTCGCGCA  CAAATTCACG  GTTGCCCTCG  tttcctaTat  gaccgtcGTC
      751 TTTTCCGCCC  TGTCTGCCGC  ATTTTCTCTg  ggcgaaagagc  ttttCtggCA
      801 GGAAATACTC  GGTATGTGCA  TCATTatccT  CAGCGGCATT  TTGAGCAGCA
      851 TCCGCCCAT  TGCCTTCAAA  CAGCGGCTGC  AAGCCCTCTT  CCGCCAAAGA
      901 TAA

```

This corresponds to the amino acid sequence [<SEQ ID 548; ORF135ng-1>] (SEQ ID NO: 548; ORF135ng-1):

```

5      1  MDTAKKDILG SGWMLVAAAC FTVMNVLIKE ASAKFALGSG ELVFWRMFLS
      51  TVTLGAAAVL RRDTRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV
     101  TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE
     151  PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT
     201  LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
     251  FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR
    301  *
```

ORF135ng-1 (SEQ ID NO: 548) and ORF135-1 (SEQ ID NO: 542) show 97.0% identity in 300 aa overlap:

```

15  orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLIKEASAKFALGSGELVFWRMFLSTVTLGAAAVL
      |||:|||||
  orf135-1        MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

  orf135ng-1.pep RRDTRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSYTSSIFLAVFSFLILKE
      |||:|||||
  orf135-1        RRDXFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE

20  orf135ng-1.pep RISVYTQAVLLLGFGAGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG
      |||:|||||
  orf135-1        RISVYTQAVLLLGFGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG

  orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
      |||:|||||
25  orf135-1        WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT

  orf135ng-1.pep VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR
      |||:|||||
  orf135-1        VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
```

30 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 66

The following DNA sequence was identified in *N.meningitidis* [<SEQ ID 549>] (SEQ ID NO: 549):

```

35      1  ATGAAGCGGC GTATAGCCGT CTTCTCCTG TCCCCGAGA TAATCCGAGT
      51  TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA
     101  TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT
     151  CTGCCCGGGA TCGCCGAAAT CGATTCCCCA TGCGGCATCG TGTTCCGGTGC
40    201  GCTCTCTTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG
```

5
10

```

251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG
301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTcAGTTCG CCsGGTTCAT
351 TGTTcAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC
401 CACATATGTT CGCAAATTTc GCCGTCTTcG CCGTCTTGGa AAAAAGGGAC
451 TTTGACCATG GCAAAATCCA AGGCGGAAAT AATGCGGCGG CGTTCcCAAA
501 AAAGcTCGCG CCAAAATAT TTGAATGTTT TACGGGCGCG TCGTCGGCA
551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC
601 CATCATATCT GCTCCTCAAC GTGTACGGTA TCTGTTTGCA CCTTACTGCG
651 GCTTTCTgcC kTCGGCATCC GATTcGGATT TGAAAAGTTC mmrwyATTcG
701 GAATAG

```

This corresponds to the amino acid sequence [<SEQ ID 550; ORF136>] (SEQ ID NO: 550; ORF136):

15

```

1 MKRRIAVFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHHQQY
51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVNR
101 NANAFALFDI GQFAXFIVQH TVNIKTvKIN IVDPHMFANF AVFAVLEKRD
151 FDHGKIQGGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA
201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE*

```

20 Further work revealed the complete nucleotide sequence [<SEQ ID 551>] (SEQ ID NO: 551):

25
30
35

```

1 ATGATGAAGC GGCgtATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG
51 AGTTTtGGGA CAACTGTTCG CGAAAATCGT CAATACAGTT CCGGCACATC
101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
151 TATCTGCCCC GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
201 TGCGCTCCTC TTCCGTcATC TGCCCGCGCA TTGCTGTAT GGTAAAGCCG
251 CCGTAGGGGA TGCCGTTCGA CACGAACATC CAGTCGCTGA TGTCGTCAAC
301 CGGAACGCAA ACGTtTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT
351 CATTGTTCAG CACACCGTAA ATATAAAGAC CGTCAAAATA AATATCGTCG
401 ATCCACATAT GTTCGCAAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG
451 GACTTTGACC ATGGCAAAAT CCAAGCGGGA AATAATGCGG CGGCGTTCCC
501 AAAAAAGCTC GCGCCAAAAA TATTGAATG TTTTACGGGC GCGTTCGTcG
551 GCACGGTTTA CCGGTTCGTC TGCCTGTTCT ACATAATAAA TGACGGAATC
601 GCCCATCATT CTGCTCCTCA ACGGTACGG TATCTGTTTG CACCTTACTG
651 CGGCTTCTG CCTTCGCAT CCGATTcGGA TTTGAAAAGT TCCAAATATT
701 CGGAATAG

```

This corresponds to the amino acid sequence [<SEQ ID 552; ORF136-1>] (SEQ ID NO: 552; ORF136-1):

40

```

1 MMKRRIAVFV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHHQY
51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAGDAVA HEHPVADVNR
101 RNANAFALFD IGQFAGFIVQ HTVNIKTvKI NIVDPHMFAN FAVFAVLEKR
151 DFDHGKIQGG NNAAAFPKKL APKIFECFTG AFVGTvYRFV CLFYIINDGI
201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

```

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 (SEQ ID NO: 550) shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) (SEQ ID NO: 554) from strain A of *N. meningitidis*:

```

5      10      20      30      40      50      59
orf136.pep  MKRRIAVFVLPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS
          |||||:| ||:||||| ||||| ||||| ||||| ||||| ||||| |||||
orf136a     MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFFIHQYLPGLAEIDS
          10      20      30      40      50      60

10     60      70      80      90     100     110     119
orf136.pep  PCGIVFGALLFRHLPAHCLYLGKAAVGDAVAHEHPVADVNNRANAFALFDIGQFAXFIVQ
          |||||:|||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf136a     PCGIVFGTLLFRHXSTHCLYLGKAAVGNAVAHEHPVADVNNRANAFALFDIGQFAGFIVQ
          70      80      90     100     110     120

15     120     130     140     150     160     170     179
orf136.pep  HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG
          |:|:||||| ||||| ||||| :|: | :|: | :|: | :|: | :|: |
orf136a     HAINVKTVKINIVDPHMFANFAXFAVLEKRALTMAKSKXXXMRRRSQKSSRQKYLNVLRA
          130     140     150     160     170     180

20     180     190     200     210     220     230
orf136.pep  AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVRYLFAPYCGFLPSASDSLKSSXXSEX
          : ||: | : : : ||||| ||||| ||||| ||||| ||||| |||||
orf136a     R---SPARFTGLSACSTXXMTESPIISAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX
          190     200     210     220     230

```

The complete length ORF136a nucleotide sequence [<SEQ ID 553>] (SEQ ID NO: 553) is:

```

25      1  ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
51      51  GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
101     101  GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
151     151  TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
201     201  TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAGCCG
30      251  CCGTAGGGAA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC
301     301  CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT
351     351  CATTGTTTCA CACGCCATAA ATGTAAGAC CGTCAAAATA AATATCGTCG
401     401  ATCCACATAT GTTCGCAAAAT TTCGCCNTCT TCGCCGTCTT GGAAAAAAGG
451     451  GCTTTGACCA TGGCAAAATC TAAGGNGNNA NNGATGCGGC GCGGTTCCCA
35      501  AAAAAGCTCG CGCCAAAAAT ATTTGAATGT TTTGCGGGCG CGTTCGCCGG
551     551  CACGGTTTAC CGGTTTGTCT GCCTGTTCTA CATAATAAAT GACGGAATCG
601     601  CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG
651     651  CGGCTTCTG CCTTCGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT
701     701  CGGAATAG

```

40 This encodes a protein having amino acid sequence [<SEQ ID 554>] (SEQ ID NO: 554):

```

45      1  MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFFIHQ
51      51  YLPGLAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNAVA HEHPVADVNN
101     101  RNANAFALFD IGQFAGFIVQ HAINVKTVKI NIVDPHMFAN FAXFAVLEKR
151     151  ALTMAKSKXX XMRRRSQKSS RQKYLNVLRA RSPARFTGLS ACST**MTES
45      201  PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

```

ORF136a (SEQ ID NO: 554) and ORF136-1 (SEQ ID NO: 552) show 73.1% identity in 238 aa overlap:

-415-

		10	20	30	40	50	60
	orf136a.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFFIHQYLPGLAEIDS					
	orf136-1	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS					
5		10	20	30	40	50	60
	orf136a.pep	PCGIVFGTLLFRHXSTHCLYGKAAVGNVAHEHPVADVNNRANAFALFDIGQFAGFIVQ					
	orf136-1	PCGIVFGALLFRHLPACLYGKAAVGDAVAHEHPVADVNNRANAFALFDIGQFAGFIVQ					
10		70	80	90	100	110	120
	orf136a.pep	PCGIVFGTLLFRHXSTHCLYGKAAVGNVAHEHPVADVNNRANAFALFDIGQFAGFIVQ					
	orf136-1	PCGIVFGALLFRHLPACLYGKAAVGDAVAHEHPVADVNNRANAFALFDIGQFAGFIVQ					
		70	80	90	100	110	120
	orf136a.pep	HAINVKTIVKINIVDPHMFANFAVFAVLEKRALTMAKSKXXMRRRSQKSSRQKYNLNLRA					
	orf136-1	HTVNIKTIVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKIFECFTG					
15		130	140	150	160	170	180
	orf136a.pep	HAINVKTIVKINIVDPHMFANFAVFAVLEKRALTMAKSKXXMRRRSQKSSRQKYNLNLRA					
	orf136-1	HTVNIKTIVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKIFECFTG					
		130	140	150	160	170	180
	orf136a.pep	R---SPARFTGLSACSTXXMTESPIISAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX					
	orf136-1	AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX					
20		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 (SEQ ID NO: 550) shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) (SEQ ID NO: 556) from *N.gonorrhoeae*:

25	orf136.pep	MKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS	59
	orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQYLPGLAEIDS	60
	orf136.pep	PCGIVFGALLFRHLPACLYGKAAVGDAVAHEHPVADVNNRANAFALFDIGQFAGFIVQ	119
	orf136ng	PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNNRANAFALFDIGQSAGFIVQ	120
30	orf136.pep	HTVNIKTIVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKIFECFTG	179
	orf136ng	HTVNIKTIVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKVFEFTG	180
	orf136.pep	AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSLKSSXXSE	234
35	orf136ng	AFAGTVYRFVCLFYIINDGIAHTAPQVRVRYLFAPYRGFLPPASDSLKSSKYSE	235

The complete length ORF136ng nucleotide sequence [<SEQ ID 555>] (SEQ ID NO: 555) is:

	1	ATGATGAAGC	GGCGTATAGC	CGTCTTCGTC	CTGCTCATGC	AGAAAATCCG
	51	GATTTTGGGA	CAACTGTTGC	CGAAAATCGT	CAATACAGTT	CCGGCACATC
40	101	GGATGCTCTT	CCAAATTTTC	GGGATGTTCT	TTTTCTTCAT	ACACCGGCAA
	151	TACCTGCCCG	GGATCGCCGA	AATCGATTCC	CCAGGCGGTA	TCGTGTTCCG
	201	TACGCTCCTC	TTCCGTCATC	TGTCCGCGCA	TTGCCTGTAC	GGTAAAGCCG
	251	CCGTAGGGGA	TGCCGTGCA	CACGAACATC	CAGTCGCTGA	TGTCGCCAAC
	301	CGGAACGCAA	ACGCTTTCGC	CTTGTTTCGAC	ATTGGTCAGT	CCGCCGGGTT
45	351	CATTGTTTCA	CACACCGTAA	ATATAAGAGC	CGTCAAAATA	AATATCGTCG

5
401 ATCCACATAT GTTCGCAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG
451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC
501 AAAAAAGCTC GCGCCAAAAG TATTGAATG TTTTACGGG GCGTTGCGCG
551 GCACGGTTTA CCGGTTTCGTC TGCCTGTTCT ACATAATAAA TGACGGAATC
601 GCCCATCATA CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACCG
651 CGGTTTCTA CCTCCGGCAT CCGATTCGGA TTTGAAAAGT TCCAAATATT
701 CGGAATAG

This encodes a protein having amino acid sequence [<SEQ ID 556>] (SEQ ID NO: 556):

10
1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQIF GMFFFFIHRQ
51 YLPGIAEIDS PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEHPVADVAN
101 RNANAFALFD IGQSAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
151 DFDHGKIQQG NNAAAFPKKL APKVFECFTG AFAGTVYRFV CLFYIINDGI
15 201 AHHTAPQVRV YLFAPYRGFL PPASDSDLKS SKYSE*

ORF136ng (SEQ ID NO: 556) and ORF136-1 (SEQ ID NO: 552) show 93.6% identity in 235 aa overlap:

20
orf136ng MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
orf136-1 MMKRRIAVFVLFPPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
orf136ng PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVANRNANAFALFDIGQSAGFIVQ
orf136-1 PCGIVFGALLFRHLPALHCLYGKAAVGDAVAHEHPVADVNNANAFALFDIGQFAGFIVQ
25
orf136ng HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAAFPKKLAPKVFECFTG
orf136-1 HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAAFPKKLAPKIFECFTG
orf136ng AFAGTVYRFVCLFYIINDGIAHHTAPQVRVRYLFAPYRGFLPPASDSDLKSSKYSEX
30
orf136-1 AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSDLKSSKYSEX

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 67

35 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 557>] (SEQ ID NO: 557):

40
1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
51 CGCCGCCGCG TTGCTTGCCG CC.TGCGGAC GCGGGAAAT AATGCTGTCC
101 GCAAGCCGGT GCAAACCGCC AAACCGCCG CAGTGGTCGG TTTGGCACTC
151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
201 GAAAGAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT
251 CGATTGTCCG CAACCTTTT GCATCGGGTA TGTGCCCCGA CCGCTCGAA

301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCAATGGG TTTATCAAAG GCGCAAAGCT GCAAAATTAC ATCAACCGAA
 401 AACTCCGCGG CATGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCC..

5 This corresponds to the amino acid sequence [<SEQ ID 558; ORF137>] (SEQ ID NO: 558; ORF137):

1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRL
 101 LEAEILGKTD LVDLTLSTNG FIKGAKLQNY INRKLRGMQI QQFPIKFAA..

Further work revealed the complete nucleotide sequence [<SEQ ID 559>] (SEQ ID NO: 559):

1 ATGGAATAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGAAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCG CAGTGGTCGG TTTGGCACTC
 151 GGTGGCGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 251 CGATTGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCGA CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTCAATC AGGGGAATGC
 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
 601 CCCGTCAGTG CCGCCCGCG GCAGGGGCG AATTTCTGTA TTGCCGTCGA
 651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT
 801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

This corresponds to the amino acid sequence [<SEQ ID 560; ORF137-1>] (SEQ ID NO: 560; ORF137-1):

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRL
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAQQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAARRQGA NFVIAVDISA RPGKNISQGF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKPQ VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
 301 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF137 (SEQ ID NO: 558) shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) (SEQ ID NO: 562) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf137.pep	MENMVTFSKIRPLLAIAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH					
5	orf137a	MENMVTFSKIRPLLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH					
		10	20	30	40	50	60
	orf137.pep	VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLLEAEILGKTDLDLTLSTNG					
10	orf137a	VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLLEAEILGKTDLDLTLSTSG					
		70	80	90	100	110	120
	orf137.pep	FIKGAKLQNYINRKLGRMQIQFPIKFAA					
15	orf137a	FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV					
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence [[SEQ ID 561](#)] ([SEQ ID NO: 561](#)) is:

	1	ATGGAATA	TGGTAACGTT	TTCAAAATC	AGACCGCTTT	TGGCAATCGC
	51	CGCCGCCG	TTGCTTGCCG	CCTGCGGCAC	GGCGGAAAT	AATGCTGCCC
20	101	GCAAGCCGGT	GCAAACCGCC	AAACCGCCG	CAGTGGTCGG	TTTGGCACTC
	151	GGTGGCGGCG	CATCTAAAGG	ATTTGCCCAT	GTAAGTATTA	TTAAGGTTTT
	201	GAAAGAAAC	GGTATTCCTG	TGAAGGTGGT	TACCGGCACA	TCGGCAGGTT
	251	CGATAGTCGG	CAGCCTTTTT	GCATCGGGTA	TGTCGCCCCA	CCGCCTCGAA
25	301	TTGGAAGCCG	AAATTTTAGG	TAAAACCGAT	TTGGTCGATT	TAACCTTGTC
	351	CACCACTGGT	TTTATCAAAG	GCGAAAAGCT	GCAAATAC	ATCAACCGAA
	401	AAGTCGGCGG	CAGGCGGATT	CAGCAGTTTC	CCATCAAATT	TGCCGCCGTT
	451	GCTACTGATT	TTGAAACCGG	CAAGGCCGTC	GCTTTCAATC	AAGGGAATGC
	501	CGGGCAGGCT	GTGCGCGCTT	CCGCCGCCAT	TCCCAATGTG	TTCCAACCGG
30	551	TTATCATCGG	CAGGCATACA	TATGTTGACG	GCGGTCTGTC	GCAGCCCGTG
	601	CCCGTCAGTG	CCGCCCGGCG	GCANGNNNG	NATNTCGTGA	TTGCCGTCGA
	651	TATTTCCGCC	CGTCCGAGCA	AAAACATCAG	CCAAGGCTTC	TTCTCTTATC
	701	TCGATCAGAC	GCTGAACGTA	ATGAGCGTTT	CCGCGTTGCA	AAATGAGTTG
	751	GGGCAGGCGG	ATGTGGTTAT	CAAACCGCAG	GTTTTGGATT	TGGGTGCAGT
35	801	CGGCGGATTC	GATCAGAAAA	AACGCGCCAT	CCGGTTGGGT	GAGGAGGCAG
	851	CACGTGCCGC	ATTGCCTGAA	ATCAAACGCA	AACTGGCGGC	ATACCGTTAT
	901	TGA				

This encodes a protein having amino acid sequence [[SEQ ID 562](#)] ([SEQ ID NO: 562](#)):

	1	MENMVTFSKI	RPLLAIAAAA	LLAACGTAGN	NAARKPVQTA	KPAAVVGLAL
40	51	GGGASKGFAH	VGIKVLKEN	GIPVKVVTGT	SAGSIVGSLF	ASGMSPDRL
	101	LEAEILGKTD	LDLTLSTSG	FIKGEKLQNY	INRKVGGRRI	QQFPIKFAAV
	151	ATDFETGKAV	AFNQGNAGQA	VRASAAIPNV	FQPVIIGRHT	YVDGGLSQPV
	201	PVSAARRXXX	XXVIAVDISA	RPSKNISQGF	FSYLDQTLNV	MSVSALQNEL
45	251	GQADVVIKPQ	VLDLGAVGGF	DQKKRAIRLG	EAAARAALPE	IKRKLAAARY
	301	*				

ORF137a ([SEQ ID NO: 562](#)) and ORF137-1 ([SEQ ID NO: 560](#)) show 97.3% identity in 300 aa overlap:

50	orf137a.pep	MENMVTFSKIRPLLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
	orf137-1	MENMVTFSKIRPLLAIAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH

	orf137a.pep	VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLDLTLSTSG
	orf137-1	VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLDLTLSTSG
5	orf137a.pep	FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
	orf137-1	FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
	orf137a.pep	FQPVIIGRHTYVDGGLSQPVPVSAARRXXXXXVIAVDISARPSKNISQGFFSYLDQTLNV
	orf137-1	FQPVIIGRHTYVDGGLSQPVPVSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV
10	orf137a.pep	MSVSALQNELGQADVVIKQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY
	orf137-1	MSVSALQNELGQADVVIKQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 (SEQ ID NO: 558) shows 89.9% identity over a 149aa overlap with a predicted ORF
 15 (ORF137ng) (SEQ ID NO: 564) from *N.gonorrhoeae*:

	orf137.pep	MENMVTFSKIRPLLAIAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH	60
	orf137ng	MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH	60
20	orf137.pep	VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLDLTLSTNG	120
	orf137ng	IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLDLTLSTSG	120
	orf137.pep	FIKGAKLQNYINRKLQGMQIQFPIKFAA	149
25	orf137ng	FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV	180

The complete length ORF137ng nucleotide sequence [<SEQ ID 563>] (SEQ ID NO: 563) is:

	1	ATGGAATA	TGGTAACGTT	TTCAAAATC	AGATCATTTT	TGGCAATCGC
	51	CGCCGCCGCG	TTGCTTGCCG	CCTGCGGTAC	GGCGGGAAC	AATGCCGCCC
30	101	GCAAGCCGGT	GCAAACCGCC	AAACCCGCG	CAGTGGTCGC	TTTGGCACTC
	151	GGTGGCGGCG	CATCTAAAGG	ATTTGCCCAT	ATAGGAATTG	TTAAGGTTTT
	201	GAAAGAAAC	GGTATTCCTG	TGAAGGTGGT	TACCGGCACA	TCGGCAGGTT
	251	CGATAGTCGG	CAGCCTTTTG	GCATCGGGTA	TGTCGCCCGA	CCGCCTCGAA
	301	TTGGAAGCCG	AGATTTTAGG	TAAAACCGAT	TTAGTCGATT	TAACCTTGTC
	351	CACCACTGGT	TTTATCAAAG	GCGAAAAGCT	GCAAAATTAC	ATCAACCGAA
35	401	AAGTCGGCGG	CAGGCAGATT	CAGCAGTTTC	CCATCAAATT	TGCCGCCGTT
	451	GCCACTGATT	TTGAAACCGG	CAAGGCCGTC	GCTTTCAATC	AAGGGAATGC
	501	CGGGCAGGCG	GTTTCGTGCTT	CCGCCGCCAT	TCCCAATGTG	TTCCAGCCAG
	551	TCATCATCGG	CAGGCACAAA	TATGTTGACG	GCGGTCTGTC	GCAGCCCGTG
	601	CCCGTCAGTG	CCGCTCGGCG	GCAGGGGGCG	AATTTCTGTA	TTGCCGTCGA
40	651	TATTTCCGCA	CGTCCGAGCA	AAAATGTCGG	TCAAGGTTTC	TTCTCTTATC
	701	TCGATCAGAC	GCTGAACGTG	ATGAGCGTTT	CCGTGTTGCA	AAACGAGTTG
	751	gggcAGGCGG	ATGTGGTTAT	CAAACCGCag	gtTTTGGATT	TGGGTGCAGT
	801	CGGCGGATTC	GATCAGAAAA	AGCGCGCCAT	CCGGTTGGGC	GAGGAGGCAG
	851	CACGTGCCGC	ATTGCCTGAA	ATCAAACGCA	AACTGGCGGC	ATACCGTTAT
45	901	TGA				

This encodes a protein having amino acid sequence [<SEQ ID 564>] (SEQ ID NO: 564):

```

1  MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL
51  GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSLL ASGMSPDRLE
101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
151 ATDFETGKAV AFNQGAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV
201 PVSAARRQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL
251 GQADVVIKPQ VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY
301 *
```

10 ORF137ng (SEQ ID NO: 564) and ORF137-1 (SEQ ID NO: 560) show 96.0% identity in 300 aa overlap:

```

      orf137ng      MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH
      orf137-1      MENMVTFSKIRPLLAIAAAALLAACGTAGNNAVRKPVQTAKPAAVGLALGGGASKGFAH

15      orf137ng      IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLVDLTLSTSG
      orf137-1      VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG

      orf137ng      FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGAGQAVRASAAIPNV
20      orf137-1      FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGAGQAVRASAAIPNV

      orf137ng      FQPVIIGRHKYVDGGLSQPVPVSAARRQGANFVIAVDISARPSKNVGQGGFFSYLDQTLNV
      orf137-1      FQPVIIGRHTYVDGGLSQPVPVSAARRQGANFVIAVDISARPGKNISQGGFFSYLDQTLNV

25      orf137ng      MSVSVLQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
      orf137      MSVSALQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 565>] (SEQ ID NO: 565):

```

35      1  ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
      51  CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGcTG CCGCTTTCCT
      101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
      151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CGGGTTTGAA
40      201 CCCCAGCCCC AAAACGGTCA AAGCCGTTT TGCGGAAACG GCAAAAGGCG
      251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
```

301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA
351 ACACGAAGGG CTGCTATTC..

This corresponds to the amino acid sequence [<SEQ ID 566; ORF138>] (SEQ ID NO: 566;

5 ORF138):

1 MFRLQFRLFP PLRTAMHILL TALLKCLSLP PLSCHLTLGN RLGLHAFYLL
51 KEDRARIVAX MRQAGLNPD P KTVKAVFAET AKGGLELAPA FFRKPEDIET
101 MFKAVHGWEH VOQALDKHEG LLF

10 Further work revealed the complete nucleotide sequence [<SEQ ID 567>] (SEQ ID NO: 567):

	1	ATGTTTCGTT	TACAATTCAG	GCTGTTTCCC	CCTTTGCGAA	CCGCCATGCA
	51	CATCCTGTTG	ACCGCCCTGC	TCAAATGCCT	CTCCCTGCTG	CCGCTTTTCT
	101	GTCTGCACAC	GCTGGGAAAC	CGGCTCGGAC	ATCTGGCGTT	TTACCTTTTA
15	151	AAGGAAGACC	GCGCGCGCAT	CGTCGCCAAT	ATGCGGCAGG	CGGGTTTGAA
	201	CCCCGACCCC	AAAACGGTCA	AAGCGGTTTT	TGCGGAAACG	GCAAAAGGCG
	251	GTTTGGAACT	TGCCCCCGCG	TTTTTCAGAA	AACCGGAAGA	CATAGAAACA
	301	ATGTTCAAAG	CGGTACACGG	CTGGGAACAT	GTGCAGCAGG	CTTTGGACAA
	351	ACACGAAGGG	CTGCTATTCA	TCACGCCGCA	CATCGGCAGC	TACGATTTGG
20	401	CGGGACGCTA	CATCAGCCAG	CAGCTTCCGT	TCCCGCTGAC	CGCCATGTAC
	451	AAACCGCCGA	AAATCAAAGC	GATAGACAAA	ATCATGCAGG	CGGGCAGGGT
	501	TCGCGGCAAA	GGAAAAACCG	CGCCTACCAG	CATACAAGGG	GTCAAAACAA
	551	TCATCAAAGC	CCTGCGTTTC	GGCGAAGCAA	CCATCGTCCT	GCCCGACCAC
	601	GTCCCCTCCC	CTCAAGAAGG	CGGGGAAGGC	GTATGGGTGG	ATTTCTTCGG
25	651	CAAACCTGCC	TATACCATGA	CGCTGGCGGC	AAAATTGGCA	CACGTCAAAG
	701	GCGTGAAAAC	CCTGTTTTTC	TGCTGCGAAC	GCCTGCCTGG	CGGACAAGGT
	751	TTGATTTTGC	ACATCCGCCC	CGTCCAAGGG	GAAATTGAACG	GCACAAGAAC
	801	CCATGATGCC	GCCGTGTTC	ACCGCAATGC	CGAATATTGG	ATACGCCGTT
	851	TTCCGACGCA	GTATCTGTTT	ATGTACAACC	GCTACAAAAT	GCCGTAA

30 This corresponds to the amino acid sequence [SEQ ID 568; ORF138-1] (SEQ ID NO: 568;
ORF138-1);

35

1	MFR LQFRLFP	PLRTAMHILL	TALLKCLSL	PLSCLHTLGN	RLGHLAFYLL
51	KEDRARIVAN	MRQAGLNPD	KTVKAVFAET	AKGGL ELAPA	FFRKPEDIET
101	MFKAVHGW EH	VQQALDKHEG	LLFITPHIGS	YDLGGRYISQ	QLPFP L TAMY
151	KPPKIK AIDK	IMQAGRVRGK	GKTAPTSIQG	VKQI I KALRS	GEATIVLPDH
201	VPSPEGGEG	VWVDFGKPA	YMTLAAKLA	HVKGVKTLFF	CCERLPGGQH
251	FDLHIRPVOG	ELNGDKAHDA	AVFN RNAEYW	IRRFPTQYLF	MYNRYKMP*

Computer analysis of this amino acid sequence gave the following results:

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 (SEQ ID NO: 566) shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) (SEQ ID NO: 570) from strain A of *N. meningitidis*:

45 orf138.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX

	orf138a	MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
		10 20 30 40 50 60
5	orf138.pep	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
	orf138a	MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
		70 80 90 100 110 120
10	orf138.pep	LLF
	orf138a	LLFITPHIGSYDLGGRIISQQLPFPLTAMYKPPKIKAIKDIMQAGRVRGKGKTAPTSIQG
		130 140 150 160 170 180

The complete length ORF138a nucleotide sequence [<SEQ ID 569>] (SEQ ID NO: 569) is:

15	1	ATGTTTCGTT	TACAATTCAG	GCTGTTTCCC	CCTTTGCGAA	CCGCCATGCA
	51	CATCCTGTTG	ACCGCCCTGC	TCAAATGCCT	CTCCCTGCTG	CCGCTTTCCT
	101	GTCTGCACAC	GCTGGGAAAC	CGGCTCGGAC	ATCTGGCGTT	TTACCTTTTA
	151	AAGGAAGACC	GCGCGCGCAT	CGTCGCCAAT	ATGCGTCAGG	CAGGCATGAA
	201	TCCCGACCCC	AAAACGGTCA	AAGCCGTTTT	TGCGGAAACG	GCAAAGGCG
20	251	GTTTGGAACT	TGCCCCCGCG	TTTTTCAGAA	AACCGGAAGA	CATAGAAACA
	301	ATGTTCAAAG	CGGTACACGG	CTGGGAACAT	GTGCAGCAGG	CTTTGGACAA
	351	ACACGAAGGG	CTGCTATTCA	TCACGCCGCA	CATCGGCAGC	TACGATTTGG
	401	GCGGACGCTA	CATCAGCCAG	CAGCTTCCGT	TCCCGCTGAC	CGCCATGTAC
	451	AAACCGCCGA	AAATCAAAGC	GATAGACAAA	ATCATGCAGG	CGGGCAGGGT
25	501	TCGCGGCAAA	GGAAAAACCG	CGCCTACCAG	CATACAAGGG	GTCAAAACAA
	551	TCATCAAAGC	CCTGCGTTCG	GGCGAAGCAA	CCATCGTCCT	GCCCGACCAC
	601	GTCCCCCTCC	CTCAAGAAGG	CGGGGAAGGC	GTATGGGTGG	ATTTCTTCGG
	651	CAAACCTGCC	TATACCATGA	CGCTGGCGGC	AAAATTGGCA	CACGTCAAAG
	701	GCGTGAAAAC	CCTGTTTTTC	TGCTGCGAAG	GCCTGCCTGG	CGGACAAGGT
30	751	TTGATTTTGC	ACATCCGCCC	CGTCCAAGGG	GAATTGAACG	GCACAAGAGC
	801	CCATGATGCC	GCCGTGTTC A	ACCGCAATGC	CGAATATTGG	ATACGCCGTT
	851	TTCCGACGCA	GTATCTGTTT	ATGTACAACC	GCTACAAAAT	GCCGTAA

This encodes a protein having amino acid sequence [<SEQ ID 570>] (SEQ ID NO: 570):

35	1	MFR LQFRLFP	PLRTAMHILL	TALLKCLSL	PLSCLHTLGN	RLGLHAFYLL
	51	KEDRARIVAN	MRQAGLNPD	KTVKAVFAET	AKGGLLELAPA	FFRKPEDIET
	101	MFKAVHGWHEH	VQQALDKHEG	LLFITPHIGS	YDLGGYISQ	QLPFPLTAMY
	151	KPPKIKAIDK	IMQAGRVRGK	GKTAPTSIQ	VKQIIKALRS	GEATIVLPDH
40	201	VPSPEGGEG	VWVDFGKPA	YTMTLAAKLA	HVKGVKTLFF	CERLPGGQ
	251	FDLHIRPVOG	ELNGDKAHDA	AVFNRNAEYW	IRRPQTQYLF	MYNRYKMP*

ORF138a (SEQ ID NO: 570) and ORF138-1 (SEQ ID NO: 568) show 99.7% identity over a 298aa overlap:

[illegible]

```

orf138a.pep  LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG
orf138-1      LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG

orf138a.pep  VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
orf138-1      VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF

orf138a.pep  CCERLPGGQGFDLHIRPVQELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
orf138-1      CCERLPGGQGFDLHIRPVQELNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP

```

10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 (SEQ ID NO: 566) shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) (SEQ ID NO: 572) from *N.gonorrhoeae*:

```

orf138.pep  MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX  60
orf138ng     MFRLQFRLFPPLRTAMHILLTALLKCLSLSLSCLHTLGNRLGHLAFYLLKEDRARIVAN  60

orf138.pep  MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQALDKHEG  120
orf138ng     MRQAGLNPDQTQTVKAVFAETAKCGLELAPAFFKKPEDIETMFKAVHGWEHVQALDKGEG  120

orf138.pep  LLF  123
orf138ng     LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG  180

```

The complete length ORF138ng nucleotide sequence [<SEQ ID 571>] (SEQ ID NO: 571) is:

```

1  ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
25  51  CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT
    101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
    151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
    201 CCCCACACG CAGACGGTCA AAGCCGTTT TGCGGAAACG GCAAAATGCG
    251 GTTTGGAAC TCCCCCGCG TTTTTCAAA AACCGGAAGA CATCGAAACA
30  301 ATGTTCAAAG CGGTACACG CTGGGAACAC GTGCAGCAGG CTTTGGAACA
    351 GGGCGAAGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGG
    401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC
    451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT
    501 GCGCGGCAAA GGCAAAACcg cgcccaccgg catACAAGGG GTCAAACAAA
35  551 tcatcaAGGC CCTGCGCGCG GCGGAGGCAA CCAAtATCCT GCCCGACCAC
    601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA
    651 ACCTGCATAc acCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG
    701 TGA AACCCCT GTTTTCTGC TGCGAACGCC TGCCGACGG ACAAGGCTTC
    751 GTGTTGCACA TCCGCCCGCT CCAAGGGGAA TTGAACGGCA ACAAGCCCA
40  801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTC
    851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAACGCC GTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 572):

5
 1 MFR LQFRLFP PLRTAMHILL TALLKCLSL SLSCLHTLGN RLGH LAFYLL
 51 KEDRARI VAN MRQAGLNPDT QTVKAVFAET AKCGLELAPA FFKKPEDIET
 101 MFKA VHGEH VQ QALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY
 151 KPPKIKAI DK IMQAGVRGK GKTAPTGIQG VKQIIKALRA GEATIILPDH
 201 VPSPQEGGGV WADFFGKPAY TMTLA AKLAH VKGVKTLFFC CERLPDGGQGF
 251 VLHIRPVQGE LNGNKAHDAA VFNRNTEYWI RRFPTQYLFM YNRYKTP*

ORF138ng (SEQ ID NO: 572) and ORF138-1 (SEQ ID NO: 568) show 94.3% identity over 299aa overlap:

10 orf138-1.pep MFR LQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGH LAFYLLKEDRARI VAN
 orf138ng MFR LQFRLFPPLRTAMHILLTALLKCLSL SLSCLHTLGNRLGH LAFYLLKEDRARI VAN
 15 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPED IETMFKA VHGEHVQ QALDKHEG
 orf138ng MRQAGLNPDTQTVKAVFAETAKCGLELAPAFFKKPED IETMFKA VHGEHVQ QALDKGEG
 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAI DKIMQAGVRGKGKTAPT SIQG
 orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAI DKIMQAGVRGKGKTAPTGIQG
 20 orf138-1.pep VKQIIKALRSGEATI VLPDHVPSPQEGGEGVWVDFFGKPAYTMTLA AKLAHVKG VKTLFF
 orf138ng VKQIIKALRAGEATIILPDHVPSPQEGG-GVWADFFGKPAYTMTLA AKLAHVKG VKTLFF
 orf138-1.pep CCERLPGGQGF DLHIRPVQGE LNGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
 orf138ng CCERLPDGGQGFVLHIRPVQGE LNGNKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKTP
 25

In addition, ORF138ng (SEQ ID NO: 572) is homologous to htrB protein (SEQ ID NO: 1147) from *Pseudomonas fluorescens*:

30 gnl|PID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
 Score = 80.8 bits (196), Expect = 9e-15
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
 Query: 101 MFKA VHGEHVQ QALDKGEGLLFITPHIGSYD-LGGYISQQLPFHLTAMYKPPKIKAI D 159
 + + V G E +++AL G+G++ IT H+G+++ L Y SQ P Y+PPK+KA+D
 Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVLNHFYCSQCKPI---IFYPKPKLKA VD 150
 35 Query: 160 KIMQAGVRGKGKTAPTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGGVWADFFGKPA 219
 ++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A
 Sbjct: 151 ELLRKQRVQLGNKVAASTKEGILSVI KEVRKGGQVGIPAD--PEPAESAGIFVPPFATQA 208
 Query: 220 YTMTLA AKLAHVKG VKTLFFCCERLPDGGQGF 250
 T + +F RLPDG G+
 40 Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (SEQ ID NO: 568) (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 14B). These experiments confirm that ORF138-1 (SEQ ID NO: 568) is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 573>] (SEQ ID NO: 573):

```

10      1  ..GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
      51  GCATGCGGTG TGAATACTT TGCGCTTCTC GGCGGCGGCG GTGTATGCGG
     101  CAGCGGTTTT GGGTGTGGTG TATGCGGCGC CGGCGCGGCG GTCGGCGTGG
     151  ATGCGCGGGC TGATGTTTTA GCCGTTTATG GTGTGCGCCG TTTGTGTTTC
     201  GGCGGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTTGC
15      251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACT GCCGCCGAT TACGGCAGGG CGGCGGCGGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCCCTCT
     401  TGAAACCGGC GTTGC GCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG
     451  GGCGAATTTG CCGCGACATT GTTCTGTGCG CGTCCGGAAT GGCAGACGCT
20      501  GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
     551  CGCGGGCGAT GGTGCTG..

```

This corresponds to the amino acid sequence [<SEQ ID 574; ORF139>] (SEQ ID NO: 574; ORF139):

```

25      1  ..AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGVV YAAPARRSAW
     51  MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
    101  LSAWDALPPD YGRAAGLGA NGFQTACRIT FPLLKPALRR GLTLAAATCV
    151  GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

```

30 Further work revealed the complete nucleotide sequence [<SEQ ID 575>] (SEQ ID NO: 575):

```

35      1  ATGGATGGAC GCGGTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
     51  GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
    101  ATGACGGTTT GCGGTGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAA
    151  CGTTTGGCGT GGACGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
    201  GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GGCGGCGCTG GCGTTTCCGG
    251  GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCACG
    301  TTGGTGGCGG GCGTGGGCGT GCTGCCCCTG TTCGGGGCGG ACGGGCTGTT
    351  GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT
    401  TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTGTGCAA
40      451  GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGCGGTG
    501  GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
    551  GCGGCGTGTG CTTGTCTTT CTGTATTGTT TTTCCGGGTT CGGGCTGGCG
    601  CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA

```

5 651 GTTGGTCATG TTCGAACTCG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
 701 TGGTGTGGG GGTAAACGGCG GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
 751 AGGCGCGCGG TTTCGATAA GGCGGTTTCC CCTGTGATGC CGTCGCCGCC
 801 GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTTGCGGCG GCGGTGTTGT
 851 CTGTGTGCTG CCTGTTTCCT TTGTTGGCAA TTGTTGTGAA AGCGTGGTCG
 901 GCCGCGGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCGGT
 951 GTGGAATACT TTGCGCTTCT CCGCGGCGGC GGTGTATGCG GCGGCGGTTT
 1001 TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG
 1051 CTGATGTTTT TGCCGTTTAT GGTGTCGCCG GTTTGTGTTT CCGCGGCGGT
 1101 GCTGCTGCTT TATCCGCAGT GGACGCTTTC GTTGCCGTTG CTGCTGGCGA
 1151 TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC
 1201 TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTTGGGTGC
 1251 AAACGCGTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG
 1301 CGTTGCGGCG CGGTCGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
 1351 GCGGCGACAT TGTTCCTGTC GCGTCCGAA TGGCAGACGC TGACGACTTT
 1401 GATTTATGCC TATTTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGGCGA
 1451 TGGTGCTGAC ATTGCTGTTG GCGGCGTTTC CGCTGGGTAT TTCTCTGCTG
 1501 TTGGACGGCG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA

20 This corresponds to the amino acid sequence [<SEQ ID 576; ORF139-1>] (SEQ ID NO: 576;
ORF139-1):

25 1 MDGRRVWVG AFALLPSAFL AVMVVAFLWA VAYDGLAWR AVLSDAYMLK
 51 RLAWTVFQAA ATCVLVPLG VPAVAVLARL AFPGRALVLR LLMLPFVMP
 101 LVAGVGV LAL FGADLLWRG RQDTPYLLLY GNVFFNLVPL VRAAYQGFVQ
 151 VPAARLQTAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
 201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVTA AAGLLYAWFG
 251 RRAVSDKAVS PVMPSPQSV GEYVLLAFAA AVLSVCCLFP LLAIIVKAW
 301 AGESWRV LME SETWQAVWNT LRFSAAAVYA AAVLGVVYAA AARRSAWMRG
 351 LMFLPFMVSP VCVSAGVLLL YPQWTASLPL LLAMYALLAY PFVAKDVLSA
 401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
 451 AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL AAFALGIFLL
 501 LDGGEKGQT ETL*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF139 (SEQ ID NO: 574) shows 94.7% identity over a 189aa overlap with an ORF (ORF139a)
 (SEQ ID NO: 578) from strain A of *N. meningitidis*:

40 orf139.pep AWSAGESWRV LMESETW HAVWNTLRFSAAA
 orf139a QSVGEYVLLAF AA VXSVCCLFXLLAI VVKAWSAGESWRV LMESETW QAVWNTXRFSA
 270 280 290 300 310 320

45 orf139.pep VYAAAVLG VVYAA PARRSAWM RGLMFX PFMVSPVC VSAGVLLL YPQWTASL PLLLAMYAL
 orf139a VYAAAVLG VVYAA AARRSAWM RGLMFL PFMVSPVC VSAGVLLL XPQWTASL PLLLAMYAL
 330 340 350 360 370 380

100 110 120 130 140 150

-427-

```

orf139.pep  LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV
              |||||
orf139a     LAYPFVAKDVLSAXDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV
              390      400      410      420      430      440

5
              160      170      180      189
orf139.pep  GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL
              |||||
orf139a     GEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNYARAMVLTLLLAALGXFLLLDGGEGG
              450      460      470      480      490      500

```

The complete length ORF139a nucleotide sequence [SEQ ID 577] (SEQ ID NO: 577) is:

```

1  ATGGATGGAC GCGGTTGGGC GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
51 GGCTTTTGTG GCGGCAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
101 ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA
15 151 CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
201 GCCTTTGGGC GTGCCTGTCT CGTGGGTGCT GGCGCGGCTG GCGTTTCCGG
251 GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCACG
301 TTGGTGGCGG GCGTGGGCGT GCTGGCTCTG TTCGGGGCGG ACGGCCTGTN
351 GTGGCGCGGC TGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT
20 401 TTTTNNACCT TCCTGTGTTG GTCAGGGCGG CATATCAGGG GTTTGTGCAA
451 GTGCCTGCGG CACGGCTTCA GACGGCACNG ACATTGGGCG CGGGGCGGTG
501 GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
551 GCGGCGGTGT CCTTGTCTTC CTGTATTGTT TTTGCGGGTT CGGGCTGGCA
601 TTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
25 651 GTTGGTCATG TTCGAACTCG ATATGCGGGT TGCTTCGGTG CTNGTGTGGC
701 TGGTGTNNGG GGTAAACNGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
751 AGGCGCGCGG TTTCCGATAA GGCNGTTTCC CCTGTGATGC CGTCGCCGCG
801 GCAGTCGGTC GGGGAATATG TGCTNCTGGC GTTTGCGGCG GCGGTGTNGT
851 CTGTGTGCTG CCTGTTCNTT TTGTTGGCAA TTGTTGTGAA AGCGTGGTGC
30 901 GCCGCGGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCGGT
951 GTGGAATACT NTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT
1001 TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG
1051 CTGATGTTTT TGCCGTTTAT GGTGTCGCCG GTTTGTGTTT CGGCGGGCGT
1101 GCTGCTGCTT NATCCGAGT GGACGGCTTC GTTGCCGCTG CTGCTGGCGA
35 1151 TGTATGCGCT GCTGGCGTAT CCGTTGTGG CAAAAGATGT TTTATCAGCC
1201 TNGATGCAC TGCCGCCGGA TTACGCGAGG GCGGCGGCGG GTTTGGGTGC
1251 AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG
1301 CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
1351 GCGGCAACCT TGTTCTNTGT GCGTCNCGAG TGGCAGACGC TGACGACTTT
40 1401 GATTTATGCC TATNTGGGAC GCGCGGGTGA NGATAATTAC GCGCGGGCGA
1451 TGGTGTGAC ATTGCTGTTG GCGGCGTTTC CGCTGGGTAT NTTCTGCTG
1501 TTGACGGCG GCGAAGGCGG AAAACGGACG GAAACGTTAT AA

```

This encodes a protein having amino acid sequence [SEQ ID 578] (SEQ ID NO: 578):

```

1  MDGRRWAVWG AFALLPSAFL AAMVVAPLWA VAAYDGLAWR AVLSDAYMLK
51  RLAWTVFQAA ATCVLVPLG VPVAWVLARL AFPGRALVLR LLMLPFVMP
101 LVAGVGVLAL FGADGLXWRG WQDTPYLLLY GNVFFXLPVL VRAAYQGFVQ
151 VPAARLQTAX TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLXGVTA AAGLLYAWFG
50 251 RRAVSDKAVS PVMPPQSV GEYVLLAFAA AVXSVCCFLX LLAIVVKAWS
301 AGESWRVLME SETWQAVWNT XRFSAAYYA AAVLVGVYAA AARRSAWMRG
351 LMFLPFMVSP VCVSAGVLLL XPQWTASLPL LLAMYALLAY PFVAKDVLSA
401 XDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
451 AATLFXSRXE WQTLTTLIYA YXGRAGXDNY ARAMVLTLLL AAFALGXFL
55 501 LDGEGGKRT ETL*

```

ORF139a (SEQ ID NO: 578) and ORF139-1 (SEQ ID NO: 576) show 96.5% homology over a 514aa overlap:

```

5  orf139a.pep  MDGRRWAVWGAFALLPSAFLAAMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
   orf139-1    MDGRRWVWVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA

   orf139a.pep  ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLXWRG
   orf139-1    ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG

10  orf139a.pep  WQDTPYLLLYGNVFFXLPVLVRAAYQGFFVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP
   orf139-1    RQDTPYLLLYGNVFFNLPLVLVRAAYQGFFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP

   orf139a.pep  WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAVASVLVWLXGVTA
   orf139-1    WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAVASVLVWLVLGVTA

15  orf139a.pep  AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVXSVCCLFXLLAIVVKAWS
   orf139-1    AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVLSVCCLFPLLAIVVKAWS

   orf139a.pep  AGESWRVLMESSETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
   orf139-1    AGESWRVLMESSETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP

20  orf139a.pep  VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLSAXDALPPDYGRAAAGLGANGF
   orf139-1    VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF

   orf139a.pep  QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNY
   orf139-1    QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY

25  orf139a.pep  ARAMVLTLLLAALGXFLLLDGGEGGKRTETLX
   orf139-1    ARAMVLTLLLAALGIFLLLDGGEGGKQTETLX

```

30 Homology with a predicted ORF from *N.gonorrhoeae*

ORF139 (SEQ ID NO: 574) shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) (SEQ ID NO: 580) from *N.gonorrhoeae*:

```

35  orf139.pep  AWSAGESWRVLMESSETWHA VWN TLRFSAAA 30
   orf139ng    QSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWSAGESRRVLMESSETWQAVWNTLRFSAAA 327

   orf139.pep  VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLPQWTASLPLLLAMYAL 90
   orf139ng    VFAAAVLGVVYAAAARL VWMRGLVFLPFMVSPVCVSAGVLLLPGW TASLPLLLAMYAL 387

```

```

orf139.pep  LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV 150
             |||||
orf139ng    LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV 447

orf139.pep  GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL 189
             |||||
5 orf139ng    GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFAVCIFLLLDNGEGG 507

```

The complete length ORF139ng nucleotide sequence [<SEQ ID 579>] (SEQ ID NO: 579) is predicted to encode a protein having amino acid sequence [<SEQ ID 580>] (SEQ ID NO: 580):

```

10      1 MDGRCWAVRG AFSLLPSAFL AVMVVAPLWA VAAYDGLAWR AVLSDAYMLK
        51 RLAWTVFQAA ATCVLVPLG VPVAWVLARL AFPGRALVLR LLMLPFVMP
       101 LVAGVGVLAL FGADGLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFAQ
       151 VPAARLQTAR TLGAGAWRPF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
       201 LLLGGSRYAT VEVEIYQLVM FELDMAGASA LVWLVLGVTA AAGLLYAWFG
15      251 RRAVSDKAVS PVMPSPQSV GEYVLLAFSV AVLSVCCCLFP LSAIVVKAWS
       301 AGESRRVLME SETWQAVWNT LRFSAAVFA AAVLGVVYAA AARRLVWMRG
       351 LVFLPFMVSP VCVSAGVLL YPGWTASLPL LLAMYALLAY PFVAKDVLSA
       401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
       451 AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL SAFAVCIFLL
20      501 LDNGEGGKRT ETL*

```

Further work revealed a variant gonococcal DNA sequence [<SEQ ID 581>] (SEQ ID NO: 581):

```

25      1 ATGGATGGAC GGTGTTGGGC GGTACGGGGT GCTTTTCCC TGCTGCCTTC
        51 GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
       101 ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAA
       151 CGTTTGGCGT GGACGGTGTT TCAGGCGGCG GCAACCTGTG TGCTGGTGCT
       201 GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GGCGCGGCTG GCGTTCCCGG
       251 GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCGTTTGT GATGCCACG
       301 CTGGTGGCGG GCGTGGGCGT GCTGGCTCTG TTCGGGGCGG ACGGGCTGTT
       351 GTGGCGCGGC CGGCAGGATA CGCCGTATCT GTTGTTGTAC GGCAATGTGT
       401 TTTTCAACCT GCCCGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGCTCAA
       451 GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGGCGTG
       501 GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
       551 GCGGCGGTG CTTGTCTTC CTGTATTGTT TTTGCGGGTT CGGGCTGCA
35      601 TTGCTGTTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
       651 GTTGTTATG TTCGAACTCG ATATGGCGGG GGCTTCGGCG CTGGTGTGGC
       701 TGGTGTGGG GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
       751 AGGCGCGCGG TTTCCGATAA GGCGGTTTCC CCCGTGATGC CGTCGCCGCC
       801 GCAATCGGTG GGGGAATATG TATTGCTGGC ATTTTCGGTG GCGGTGTTGT
       851 CCGTGTGCTG CCTGTTTCTT TTGTCGGCAA TTGTTGTGAA AGCGTGCTCG
       901 GCCGCGGAAT CGCGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCAGT
       951 GTGGAATACT ttGCGCTTTT CGGCGGCGGC GGTGTTTGGC GCGGCGGTTT
40      1001 TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGCTGGTGTG GATGCGCGGA
       1051 CTGGTGTGTT TACCGTTTAT GGTGTCGCCG GTTTGTGTTT CGGCGGGCGT
       1101 GCTGCTGCTT TATCCGGGGT GGACGGCTTC GTTACCGCTG CTGCTGGCGA
       1151 TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCGGCC
       1201 TGGGATGCAC TGCCGCGGGA TTACGGCAGG GCGGCGGCAG GTTTGGGCGC
       1251 AAACGGCTTT CAGACGGCAT GCCGTATCAC GTTCCCCCTC TTGAAACCGG
       1301 CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CGACGTGTGT GGGCGAATTT
50      1351 GCGGCAACCT TGTTCCTGTC GCGTCCGGAA TGGCAGACGT TGACGACTTT
       1401 GATTTATGCC TATTTGGGGC GTGCGGGTGA GGACAATTAT GCGCGGGCAA
       1451 TGGTGTGAC ATTGCTGTTG TCGGCATTG CGGTGTGCAT TTTCTGCTG
       1501 TTGGACAACG GCGAAGGCGg aaaACGGACG GAAACGTTAT AA

```

This corresponds to the amino acid sequence [<SEQ ID 582; ORF139ng-1>] (SEQ ID NO: 582; ORF139ng-1):

```

      1 MDGRCWAVRG AFSLLPSAFL AVMVVAPLWA VAAVDGLAWR AVLSDAYMLK
      51 RLAWTVFQAA ATCVLVPLG VPVAWVLARL AFPGRALVLR LLMLPFVMP
100    LVAGVGVLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFAQ
150    VPAARLQTAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
200    LLLGGSRYAT VEVEIYQLVM FELDMAGASA LVWLVLGVTA AAGLLYAWFG
250    RRAVSDKAVS PVMPSPQSV GEYVLLAFSV AVLSVCCLFP LSAIVVKAWS
300    AGESRRVLME SETWQAVWNT LRFSAAAVFA AAVLGVVYAA AARRLVWMRG
350    LVFLPFMVSP VCVSAGVLLL YPGWTASLPL LLAMYALLAY PFVAKDVLSA
400    WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
450    AATFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL SAFAVCIFLL
500    LDNGEGGKRT ETL*

```

15 ORF139ng-1 (SEQ ID NO: 582) and ORF139-1 (SEQ ID NO: 576) show 95.9% identity over 513aa overlap:

```

      orf139ng      MDGRCWAVRGAFSLLPSAFLAVMVVAPLWAVAAVDGLAWRAVLSDAYMLKRLAWTVFQAA
      orf139-1      MDGRRWVVGAFALLPSAFLAVMVVAPLWAVAAVDGLAWRAVLSDAYMLKRLAWTVFQAA
20    orf139ng      ATCVLVPLGVPVAWVLARLAFPGRALVLRLLMLPFVMP
      orf139-1      ATCVLVPLGVPVAWVLARLAFPGRALVLRLLMLPFVMP
      orf139ng      RQDTPYLLLYGNVFFNLPVLVRAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
25    orf139-1      RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
      orf139ng      WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAGASALVWLVLGVTA
      orf139-1      WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMASVWLVLGVTA
30    orf139ng      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
      orf139-1      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
      orf139ng      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
      orf139        AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
35    orf139ng      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF
      orf139-1      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF
      orf139ng      QTACRITFPLLPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
40    orf139-1      QTACRITFPLLPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
      orf139ng      ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL
      orf139-1      ARAMVLTLLLSAFALGIFLLLDNGEGGKRTETL

```

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 70

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 583>] (SEQ ID NO: 583):

```

1  ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
51  GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAGA TTCCGCATCC
101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
151 GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
251 AACGTTTGGT C...
```

15 This corresponds to the amino acid sequence [<SEQ ID 584; ORF140>] (SEQ ID NO: 584; ORF140):

```

1  MDGWTQTLQA QTLGLISAAA IILILILIVR FRIHALLTLV IVSLLTALAT
51  GLPTGSIVKD ILVKNFPGTL GGVALLVGLG AMLERLV..
```

20 Further work revealed the complete nucleotide sequence [<SEQ ID 585>] (SEQ ID NO: 585):

```

1  ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
51  GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
25  151 GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
251 GACGTTTGGT CGAAACATCC GGCGGCGCAC AGTCGCTGGC GGACGCGCTG
301 ATCCGGATGT TCGCGAAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
30  451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC
501 GCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
551 GCCAAGTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT
601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT
651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC
35  701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
751 ATTTTCCTGA ATACCGCGCT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
801 TGCGGACGAA ACCTGGGTTC AGACGGCAA AATAATCGGT TCGACACCGA
851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA
901 CGCGGCGAAA GCGGCAGCGC GTTGGAAAAA ACCGTGGACG GCGCACTCGC
40  951 CCCGTTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGC
1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
1051 GATTTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTTGTCT CTTGGCACT
1101 GCGTATCGCG CAAGGTTCCG CAACCGTCGC CCTGACCACC GCCGCCGCGC
1151 TGATGGCTCC TGCCGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC
45  1201 TGTATCGTAT TGGCAACGGC GGCAGGTTCTG GTCGGTTGCA GCCACTTCAA
```

1251 CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGGC
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

- 5 This corresponds to the amino acid sequence [<SEQ ID 586; ORF140-1>] (SEQ ID NO: 586; ORF140-1):

1 MDGWTQTLA QTLGGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
 51 GLPTGSIVND ILVKNFPGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
 10 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
 251 IFLNTGVSAL ISEKLVSAD E TWVQTAKIIG STPIALLISV LVALFVLGRK
 301 RGESGSALEK TVDGALAPVC SVLITGAGG MFGGVLRASG IGKALADSMA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 15 401 CIVLATAAGS VGCSEFNDSG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
 451 FALSALLFAI V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 20 ORF140 (SEQ ID NO: 584) shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) (SEQ ID NO: 588) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf140.pep		MDGWTQTLA	QTLGGISAAA	IILILILIVK	FRIHALLTLV	IVSLLTALAT	GLPTGSIVKD
25 orf140a		MDGWTQTLA	QTLGGISAAA	IILILILIVK	FRIHALLTLV	IVSLLTALAT	GLPTGSIVND
		10	20	30	40	50	60
		70	80				
orf140.pep		ILVKNFPGTL	GGVALLVGLG	AMLERLV			
30 orf140a		VLVKNFPGTL	GGVALLVGLG	AMLERLVETS	GGAQSLADAL	IRMFGEKRAP	FALGVASLIF
		70	80	90	100	110	120

The complete length ORF140a nucleotide sequence [<SEQ ID 587>] (SEQ ID NO: 587) is:

35 1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
 51 GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
 151 GGTTTGCCCA CAGGCAGCAT TGTCACGAC GTACTGGTCA AAAACTTCGG
 201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GGCGGCGCAC AGTCGCTGGC GGACGCGCTG
 40 301 ATCCGATGT TCGGGGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
 351 GCTGATTTT GGTCTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAACAGGA CGTACTGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCC
 501 GCCCCATCCG AGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
 45 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT
 601 AGCGGCTATA TGCTCGGCAA AGTGTGTTGGG CGCACCATCC ATGTTCCCGT
 651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC

5
10
15

```

701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
751 ATTTTCCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
801 TCGCGACGAA ACCTGGGTTC AGACGGCAAA AATAATCGGT TCGACACCGA
851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGCTCT GGGACGCAAA
901 CGCGGCGAAA GCGGCAGCGC GTTGGAAGAA ACCGTGGACG GCGCACTCGC
951 CCCCCGTCTGT TCCGTGATTG TGATTACCGG CGCGGGCGGT ATGTTGCGCG
1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
1051 GATTTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTTGTCG CCTTGGCACT
1101 GCGTATCGCG CAAGGTTCCG CAACCGTCG CCTGACCACC GCCGCCGCGC
1151 TGATGGCTCC TGCCGTGTCG GCCGCCGGCT TTACCGACTG GCAGCTCGCC
1201 TGTATCGTAT TGGCAACGGC GGCAGGTTCT GTCGGTTGCA GCCACTTCAA
1251 CCACTCCGGC TTCTGGCTGG TCGGCCGCTT CTTGGACATG GACGTACCGA
1301 CCACGTGAA AACCTGGACG GTCACCAAAA CCCTCATCGC ACTCATCGGC
1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

```

This encodes a protein having amino acid sequence [<SEQ ID 588>] (SEQ ID NO: 588):

20
25

```

1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
251 IFLNTGVSAL ISEKLVSAD E TWVQTAKIIG STPIALLISV LVALFVLGRK
301 RGESGSALEK TVDGALAPVC SVILITGAGG MFGGVLRASG IGKALADSM
351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
401 CIVLATAAGS VGCSEFNDG FVLVGRLLDM DVPTTLKTWT VNQTLIALIG
451 FALSALLFAI V*

```

ORF140a (SEQ ID NO: 588) and ORF140-1 (SEQ ID NO: 586) show 99.8% identity over a 461aa overlap:

30
35
40
45

```

orf140-1.pep MDGWTQTLAQTLLGISAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60
orf140a MDGWTQTLAQTLLGISAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60

orf140-1.pep ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120
orf140a VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120

orf140-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 180
orf140a GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 810

orf140-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPVPELLSGGTQDNDLPKEPAKAGTV 240
orf140a ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPVPELLSGGTQDNDLPKEPAKAGTV 240

orf140-1.pep VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300
orf140a VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300

orf140-1.pep RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC 360
orf140a RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC 360

orf140-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDG 420

```

```

orf140a      FLVALALRIAQGSATVALTTAAALMAPAVAAAAGFTDWQLACIVLATAAGSVGCSHFND SG 420

orf140-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461
              |||||
orf140a      FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV 461

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 (SEQ ID NO: 584) shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) (SEQ ID NO: 590) from *N.gonorrhoeae*:

```

10 orf140.pep  MDGWTQTLQAQTLGISAIAIILILILIVRFRIHALLTLVIVSLLTALATGLPTGSIVKD 60
    ||| |||||
orf140ng      MDGRTQTLQAQTLGISAIAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND 60

orf140.pep    ILVKNFGGTLGGVALLVGLGAMLERLV 87
              :|||
orf140ng      VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF 120

```

15 The complete length ORF140ng nucleotide sequence [<SEQ ID 589>] (SEQ ID NO: 589) was predicted to encode a protein having amino acid sequence [<SEQ ID 590>] (SEQ ID NO: 590):

```

20 1 MDGRTQTLQA QTLGISAIA IILILILIVK FRIRALLTLV IASLLTALAT
51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
101 IRMFGEKRAP FAPGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
151 FALASVGAFS VMHVFLLPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
201 SGYMLGKVLG RAIHVPVPEL LSGGTQSDSP PKEPAKAGTV VAVMLIPMLL
251 IFLNTGVSAL ISEKLVSAD E TWVQTAKMIG STPVALLISV LAALLVLGRK
301 RGESGSTLEK TVDGALAPAC SVLITGAGG MFGGVLRASG IGKALADSMA
351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
25 401 CIVLATAAGS VGCSEHFND SG FWLVGRLLSDM DVPTTLKTWT VNQTLIAFIG
451 FALSALLFAI V*

```

Further work revealed a variant gonococcal DNA sequence [<SEQ ID 591>] (SEQ ID NO: 591):

```

30 1 ATGGACGGCC GGACACAGAC GCTGTCCGCG CAAACCTTGT TGGGCATTTC
51 GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
101 GCGCGCTGCT GACACTGGTC ATCGCCAGCC TGCTGACGGC TTTGGCAACC
151 GGTTTGCCCA CAGGCAGCAT CGTCAACGAC GTACTGGTCA AAAACTTCGG
201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGTCTGGGC GCAATGCTCG
35 251 GACGTTTGGT AGAAACATCC GCGGCGCAC AGTCGCTGGC GGACGCGCTG
301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGTCCGGC GCGTTGCCTC
351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
401 TGCCCATCGT ATTCGCCACC GCACGGCGCA TGAACAGGA CGTACTGCCC
451 TTCGCGCTTG CCTCCGTCGG CGCATTTTCC GTCATGCACG TCTTCTGCC
501 GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
40 551 GCCAGGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT
601 AGCGGCTATA TGCTCGCAA AGTGTGTTGGG CGCGCCATCC ATGTTCCCGT
651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAGCGACCCG CCGAAAGAAC
701 CTGCCAAAGC AGGAACGGTC GTCGCCGTCA TGCTGATTCC CATGCTGCTG
751 ATTTTCTCTGA ATACGGCGT ATCAGCCCTC ATCAGCGAAA AACTCGTAAG
45 801 TGCGGACGAA ACTTGGGTTT AGACGGCAAA AATGATCGGT TCGACACCTG
851 TCGCCCTTCT GATTTCCGTA TTGGCCGCAC TGTGGTCTT GGGACGCAAA
901 CGCGGCGAAA GCGGCAGCAC GTTGGAAAAA ACCGTGGACG GCGCACTCGC

```


-436-

```

orf140ng-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIAFIGFALSALLFAIV
                |||||||||||||||||||||:|||||||||||||
orf140-1        FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV

```

5 Furthermore, ORF140ng-1 (SEQ ID NO: 592) is homologous to an *E.coli* protein (SEQ ID NO: 1148):

```

10 gi|882633 (U29579) ORF_o454 [Escherichia coli] )gi|1789097 (AE000358) o454;
    This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
    protein Gntp_BACLI SW: P46832 [Escherichia coli] Length = 454
    Score = 210 bits (529), Expect = 1e-53
    Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

    Query: 88 ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
              E SGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
    Sbjct: 80 EHSAGAESLANYFSRKLGDKRTIAALTLAAFFLGIPVFFDVGFILAPIIYGFQAKVAKIS 139

15 Query: 148 VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
              L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K
    Sbjct: 140 PLKFGLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVGYFAAK 198

    Query: 208 VLGRAIHVPVPELL-----SGGTQSDSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
              ++ + + E+L G T+ SD P A V +++IP+ +I T
    Sbjct: 199 IINKRQYAMSVEVLEQMQ LAPASEEGATKLSDKINPPGVA-LVTSILIVIPAIIMAGT-- 255

    Query: 258 SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
              +S L+ + T ++IGS +RG S + AL
    Sbjct: 256 ---VSATLMPPSHPLLGLTLQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312

25 Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377
              A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
    Sbjct: 313 TAAVVILVTGAGGVFGKVLVESGVGKALANMLQ MIDLPLLPAAFIISLALRASQGS--AT 370

    Query: 378 XXXXXXXXXXXXXXXGFTDQWLACIVLATAAGSVGCSHFNDSGFWLVGRLLDMDVPTTLK 437
              G Q + LA G +G SH NDSGFW+V + L + V LK
    Sbjct: 371 VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430

30 Query: 438 TWTVNQTLIAFIGFALSALLFAIV 461
              TWTV T++ F GF ++ ++A++
    Sbjct: 431 TWTVLTTILGFTGFLITWCVWAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 71

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 593] (SEQ ID NO: 593):

```

1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
51  TTTGCTGTCG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
101 GCGTATTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
151 AACTTTTTTG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
201 CTGTATCGGG CTGATTCCAG TTGCCCATT CTCAACCCC GCTGCCGCCG
251 CCTTGCCGC CGCCGGAAGT GTGCTGCACG GTTATTCTTT GGCTCGCCGG
301 CGCGTGATTG CCGCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
351 GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCG
401 TACTGATGTT TTTCCGTCCG ..

```

This corresponds to the amino acid sequence [[SEQ ID 594](#); [ORF141](#)>] ([SEQ ID NO: 594](#); [ORF141](#)):

```

1  ..DFGISPVYLW VAAAFKHLIS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAAFAAAGL VLHGYSLARR
101 RVIAASFLLG TGWTLMSLAA AYPAAFALML PLPVLMFRRP ..

```

Further work revealed the complete nucleotide sequence [[SEQ ID 595](#)] ([SEQ ID NO: 595](#)):

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCCTGGTT TGGCCCGGCG
101 TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCC TTGGTTGCCC ATCTGTTTCGG
201 TCAAACCGAT TTCGGCATA CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
251 TCAAACATTT GCTGTCGCCG TGGGCTGCCG ACTCATACGA TGCCGCACGC
301 TTTGCAGGCG TATTTTTCG CGTTATCGGA CTGACTTCCT GCGGCTTTGC
351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
401 TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTTCCT CAACCCCGCT
451 GCCGCCGCCT TTGCCGCCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
501 TCGCCGGCGC GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
551 TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTGCCCCT GATGCTGCCC
601 TTGCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTGATGTT
651 GACGGCAGTC GCCTCACTTG CCTTGCCCT GCCGCTTATG ACCGTTTACC
701 CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTGCGCA ATGGCTCGAC
751 TATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACGTT AGACGGCATT
801 CAGTTTGTTC TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
851 TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
901 TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
951 CGTCAATCCG CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCCGCCGC
1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
1051 GCGTTTGTC ACTGGTTCGG CATTATGGCG TTCGACTGT TTGCCGTGTT
1101 CCTGTGGACG GGCTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
1151 CCGAACGCGC GCGCTATTT AGCCCGTATT ATGTTCTGA TATCGATCCC
1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGG TGTGGGCGAT
1251 TACCCGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
1301 GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
1351 GACGCGGCGA AAAGCCACGC GCCGTCGTC CGGAGTATGG AGGCATCGCT
1401 TTCCCGGAA TTGAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
1451 TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGCGACA
1501 TTGCCGACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCTCTCT
1551 GCCCAAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
1601 CGCGTCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAATCGGG
1651 GAAAATATAT AA

```

This corresponds to the amino acid sequence [[SEQ ID 596](#); [ORF141-1](#)] ([SEQ ID NO: 596](#); [ORF141-1](#)):

-438-

5
10
15
20
25
30
35
40
45
50
55

```

1  MLTYTPPDAR PPAKTHEKPW LLLLMFAFWL WPGVFSDHLW NPDEPAVYTA
51  VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAAPKHLSP WAADSYDAAR
101 FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPVAHFLNPA
151 AAFAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD
251 YHVFSGTGGV RHVQTAFSLF YYLKNLLWFA LPALPLAVWT VCRTRLFSTD
301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRRGAA
351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
401 IPMAVAVLFT PLWLWATR KIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
451 DAAKSHAPVV RSMEASLSP LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
501 LPHRVGDVQC RYRIVLLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
551 ENI*

```

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF141 (SEQ ID NO: 594) shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) (SEQ ID NO: 598) from strain A of *N. meningitidis*:

20
25
30
35

```

      orf141.pep      10      20      30
                        DFGISPVYLWVAAAFKHLLSPWAADSYDVA
      orf141a      WNPDEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAA
                    40      50      60      70      80      90

      orf141.pep      40      50      60      70      80      90
      RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAFAAAGL
      orf141a      RFAGVFFAVVGLTSCGFAGFNFLGRHHGRSVVLILIGCIGLIPVHFLNPAFAAAGL
                    100     110     120     130     140     150

      orf141.pep      100     110     120     130     140
      VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP
      orf141a      VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTA
                    160     170     180     190     200     210

      orf141a      VASLAFALPLMTVYPLLLAKTQPALFAQWLDHVFSGTGGVRHIQTAFSLFYLLKNLLWF
                    220     230     240     250     260     270

```

The complete length ORF141a nucleotide sequence [<SEQ ID 597>] (SEQ ID NO: 597) is:

40
45

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
51  AAAGCCGTGG CTGTTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
101 TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCT TTGGTTGCCC ATCTGTTCCG
201 TCAAATCGAT TTCGCGATAC CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
251 TCAAACATTT GCTGTCGCGG TGGGCTGCCG ACCCGTATGA TGCCGCACGC
301 TTTGCCGGCG TGTTTTTCGC CGTTGTCGGA CTGACTTCCT GCGGCTTTGC
351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
401 TCATCGGCTG TATCGGGCTG ATTCCGACCG TACACTTCT CAACCCGCT
451 GCCGCCGCTT TTGCCCGCG CGGACTGGTG CTGCACGGTT ATTCTTTGGC
501 TCGCCGGCGC GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGTGGACGC
551 TGATGTCGTT GGCAGCAGCT TATCCGGCGG CATTTGCCCT GATGCTGCCC

```

```

5  601 CTGCCCCGTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
    651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
    701 CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTCGCGCA ATGGCTCGAC
    751 GATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACATTC AGACGGCATT
    801 CAGTTTGT TT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCTGCGC
    851 TGCCCGTGGC GGTTTGGACG GTTTGCCGCA CGCGCTGTT TTCGACCGAC
    901 TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTTGG TGCTGCTTGC
    951 CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCGCCCGC
10 1001 TTGCCCTGTT CGGCGCGCGC CAACTGGACA GCCTGAGACG CGGCGCGGCG
    1051 GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
    1101 CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
    1151 CCGAACGCGC CGCCTATTTT AGCCCGTATT ATGTTCTGA TATCGATCCC
    1201 ATTCCGATGG CGGTTGCCGT ACTGTCACA CCCTGTGGC TGTGGGCGAT
    1251 TACCCGCAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
15 1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
    1351 GACGCGGCGA AAAGCAGCGC GCCCGTCGTC CGGAGTATGG AGGCATCGCT
    1401 TTCCCCGAA TTA AACCGG AGCTTTCAGA CGGCATCGAG TGTATCGACA
    1451 TAGGCGGCGG CGACCTACAC ACGCGGATTG TTTGGACGCA GTACGGCACA
    1501 TTGCCGACC GCGTCGCGCA TGTACAATGC CGCTACGCA TCGTCCGCTT
20 1551 GCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGGTG
    1601 CGGCCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG
    1651 GAAATATAT TAAAAACAAC AGATTGA

```

This encodes a protein having amino acid sequence [SEQ ID 598] (SEQ ID NO: 598):

```

25 1 MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFS HDLW NPDEPAVYTA
    51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAAPKHLSP WAADPYDAAR
    101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA
    151 AAFAAAGLV LHGYSLARRR VIAASFL LGT GWTLM SLAAA YPAAFALMLP
30 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD
    251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWFA LPALPLAVWT VCRLRFSTD
    301 WGILGVVWML AVLVLLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRRGAA
    351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYVVPDIDP
    401 IPMAVAVLFT PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPLW
    451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT
35 501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKTG
    551 ENILKTTD*

```

ORF141a (SEQ ID NO: 598) and ORF141-1 (SEQ ID NO: 596) show 98.2% identity in 553 aa overlap:

```

40 orf141a.pep MLTYTPPDARPPAKTHEKPWLLLMAFAWLWPGVFS HDLWNPDEPAVYTAVEALAGSPTP
    orf141-1 MLTYTPPDARPPAKTHEKPWLLLMAFAWLWPGVFS HDLWNPDEPAVYTAVEALAGSPTP

    orf141a.pep LVAHLFGQIDFGIPPVYLWVAAAPKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
    orf141-1 LVAHLFGQIDFGIPPVYLWVAAAPKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
45 orf141a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAAFAAAGLV LHGYSLARRRVIAASFL LGT
    orf141-1 FLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAAFAAAGLV LHGYSLARRRVIAASFL LGT

    orf141a.pep GWTLM SLAAAYPAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
    orf141-1 GWTLM SLAAAYPAAFALMLPLPVLMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
50

```

	orf141a.pep	QPALFAQWLDDHVFQTFGGVRHIQTAFSLFYLLKNLLWFALPALPLAVWTVCTRFLFSTD
	orf141-1	QPALFAQWLDYHVFQTFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCTRFLFSTD
5	orf141a.pep	WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
	orf141-1	WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
	orf141a.pep	FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMMAVAVLFTPLWLWAITRK
	orf141-1	FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMMAVAVLFTPLWLWAITRK
10	orf141a.pep	NIRGRQAVTNWAAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMESLSPELKRELSGDGIE
	orf141-1	NIRGRQAVTNWAAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMESLSPELKRELSGDGIE
	orf141a.pep	CIDIGGGDLHTRIVWTQYGTLPVRVGDVQCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
15	orf141-1	CIGIGGGDLHTRIVWTQYGTLPVRVGDVQCRYRIVLLPQNADAPQGWQTVWQGARPRNKD
	orf141a.pep	SKFALIRKTGENI
	orf141-1	SKFALIRKIGENI

Homology with a predicted ORF from *N.gonorrhoeae*

- 20 ORF141 (SEQ ID NO: 594) shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) (SEQ ID NO: 600) from *N.gonorrhoeae*:

	orf141.pep	DFGISPVYLWVAAAFKHLSPWAADSVDVA	30
	orf141ng	WNPAEPAVYTAVEALAGSPTPLVAHLFGQTDGFI PPVYLWVAAAFKHLSPWAAHPYDAA	126
25	orf141.pep	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAFAAAGL	90
	orf141ng	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRSVVLIHIGCIGLIPVAHFFNPAAAFAAAGL	186
	orf141.pep	VLHGYSLARRRVIAASFLLGTGWTLSLAAAYPAAFALMLPLPVLMMFFRP	140
30	orf141ng	VLHGYSLARRRVIAASFLLGTGWTLSLAAAYPAAFALMLPLPVLMMFFRPQSRRLMLTA	246

An ORF141ng nucleotide sequence [<SEQ ID 599>] (SEQ ID NO: 599) was predicted to encode a protein having amino acid sequence [<SEQ ID 600>] (SEQ ID NO: 600):

35	1	MPSEAVSARP	LCEYLLHLAI	RPFLTLMLT	YTPPDARPPA	KTHEKPWLLL
	51	LMAFAWLWPG	VFSHDLWNPA	EPAVYTAVEA	LAGSPTPLVA	HLFGQTDGFI
	101	PPVYLWVAAA	FKHLSPWAA	HPYDAARFAG	VFFAVIGLTS	CGFAGFNFLG
	151	RHHGRSVVLI	HIGCIGLIPV	AHFFNPAAAA	FAAAGLVLHG	YSLARRRVIA
	201	ASFLLGTGWT	LMSLAAAYPA	AFALMLPLPV	LMFFRPWQSR	RLMLTAVASL
	251	AFALPLMTVY	PLLLAKTQPA	LFAQWLNHYH	FGTFGGVRHI	QRAFSLFHYL
40	301	KNLLWFAPPG	LPLAVWTVCR	TRLFSTDWGI	LGIVWMLAVL	VLLAFNPQRF
	351	QDNLVWLLPP	LALFGAAQLD	SLRRGAAAFV	NWFGIMAFGL	FAVFLWTGFF
	401	AMNYGWPAKL	AERAAYFSPY	YVPDIDIPM	AVAVLFTPLW	LWAITRKNIR
	451	GRQAVTNWAA	GVTTLWALLM	TLFLPWLDAA	KSHAPVVRSM	EASFSPELKR

501 ELSDGIECIG IGGGDLHTRI VWTQYGTLPV RVGDVRCRYR IVRLPQNADA
 551 PQGWQTVWQG ARPRNKDSKF ALIRKIGENI LKTTD*

Further work revealed the following gonococcal DNA sequence [SEQ ID 601] (SEQ ID NO: 601):

1 ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
 51 AAAACCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGCTG TGGCCCGGCG
 101 TGTTTCCCA CGATTGTGG AATCTGCCG AACCTGCCGT CTATACCGCC
 151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCG TTGGTTGCCG ATCTGTTCCG
 201 TCAAACCGAT TTCGCATAC CGCCCGTGTA TCTTTGGGTT GCCGCGCAT
 251 TCAAACATT GCTGTCGCCG TGGGCAGCCG ACCCGTATGA TGCCGCACGC
 301 TTTGCAGGCG TATTTTTCG CGTTATCGGA CTGACTTCTT GCGGCTTTGC
 351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTT GTTTAATCC
 401 ATATCGGCTG TATCGGCTG ATTCCGGTTG CCCATTTCCT CAATCCcgcc
 451 gccgcccgcct tTGCCCGCGC CGGACTGGTG CTGCacgget actcgetgGC
 501 ACGCCGGCGC GTGATtgcg cctctTtccT GCTCGGTACG GGTGGACGT
 551 TGATGTCGCT GCGGCAGCT TATCCGGCGG CGTTTGCCTG GATGCTGCCC
 601 CTGCCCCTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
 651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
 701 CGCTGCTctt gGCAAAAACG CAGCCCGCGC TGTTTGCCTA ATGGCTCAAC
 751 TATCACGTTT TCGGTACGtT cggcgGCGTG CGGCacaTTC AGAggGCatT
 801 Cagtttgttt cactatctgA AAaatctgct ttggttcgca ccgcccgggC
 851 TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CACGCTGTT TCGACCGAC
 901 TGGGGGATT TGGGCATTGT CTGGATGCTT GCCGTTTGG TGCTGCTCGC
 951 CTTTAATCCG CAGCGTTTTT AAGACAACT CGTCTGGCTG CTGCCGCCCG
 1001 TTGCCCTGTT CCGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
 1051 GCTTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGGCTGT TTGCCGTGTT
 1101 CCTGTGGACG GGCTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
 1151 CCGAACGCGC GCCTACTTC AGCCCGTATT ACGTTCCCGA CATCGATCCC
 1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGG TGTGGGCGAT
 1251 TACCCGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
 1301 GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
 1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGTT
 1401 TTCCCGGAA TTAACACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
 1451 TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
 1501 TTGCCGACC GCGTCGGCGA TGTCCGTTGC CGCTACCGTA TCGTCCGCT
 1551 GCCCAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGGTG
 1601 CGCGCCCGC CAACAAAGAC AGTAAGTTTG CACTGATACG GAAAATCGGG
 1651 GAAAATATAT TAAAAACAAC AGATTGA

This corresponds to the amino acid sequence [SEQ ID 602; ORF141ng-1] (SEQ ID NO: 602; ORF141ng-1):

1 MLTYTPPDAR PPAKTHEKPW LLLLMAFAWL WPGVFSHDLW NPAEPAVYTA
 51 VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAAPKHLSP WAADPYDAAR
 101 FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLIHIGCIGL IPVAHFLNPA
 151 AAFAAAGLV LHGYSLARRR VIAASFLGT GWTLMSLAAA YPAAFALMLP
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLN
 251 YHVFGTGGV RHIQRAFSLF HYLKNLLWFA PPGLPLAVWT VCRTRLFSTD
 301 WGILGIVWML AVLVLLEFNP QRFQDNLVWL LPPLALFGAA QLDLRRGAA
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
 401 IPMAVAVLFT PLWLWAIKTR NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
 451 DAAKSHAPVV RSMEASFSPK LKRELSDGIE CIGIGGDLH TRIWWTQYGT
 501 LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
 551 ENILKTTD*

ORF141ng-1 (SEQ ID NO: 602) and ORF141-1 (SEQ ID NO: 596) show 97.5% identity in 553 aa overlap:

```

5  orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAPAVYTAVEALAGSPTP
   orf141-1      MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP

   orf141ng-1.pep LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVIGLTSCGFAGFN
   orf141-1      LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN

10  orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
   orf141-1      FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT

   orf141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
   orf141-1      GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT

15  orf141ng-1.pep QPALFAQWLNHYHVFGTGGVVRHIQRAFSLFHYLKNLLWFAPPGLPLAVWTVCRTRLFSTD
   orf141-1      QPALFAQWLDYHVFGTGGVVRHVQTAFSLFYLYLKNLLWFALPALPLAVWTVCRTRLFSTD

   orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
   orf141-1      WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA

20  orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAPKLAERAAYFSPYYVPDIDPIPMMAVAVLFTPLWLWAITRK
   orf141-1      FGLFAVFLWTGFFAMNYGWPAPKLAERAAYFSPYYVPDIDPIPMMAVAVLFTPLWLWAITRK

   orf141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASFSP ELKRELS DGIE
   orf141-1      NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASLSPEL KRELS DGIE

25  orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGQTVWQGARPRNKD
   orf141-1      CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVLLPQNADAPQGQTVWQGARPRNKD

   orf141ng-1.pep SKFALIRKIGENILKTDX
   orf141-1      SKFALIRKIGENIX
30

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 603>] (SEQ ID NO: 603):

1 ..CAATCCGCCA AATGGTTATC GGGCCAAACT CTAGTCGGCA CAGCAATTGG
51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101 CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
151 AGCGGTTTTTC AGGTAGGCTA TACGTTTTAA

This corresponds to the amino acid sequence [<SEQ ID 604; ORF142>] (SEQ ID NO: 604; ORF142):

1 ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
51 SGFQVGYTF*

Further work revealed the complete nucleotide sequence [<SEQ ID 605>] (SEQ ID NO: 605):

1 ATGGATAATT CGGGTAGTGA GCGCAGAGGA AAATACCAAG GAAATATCAC
51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTC TATGTAAATT
101 ATGGACGTTT GATTGGCGGT ACGCCCGATG AGGAAAGTTT TGACGGCCAT
151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT
201 CGGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT
301 ACTGATTTCG GCTTCAACCG CCTGTGTAT CGTGATGCCA AACGCAAAAC
351 CTATCTCGGT GTAAACTGT GGATGAGGGA AACAAAAAGT TACATTGATG
401 ATGCCGAAC TACTGTACAA CCGCGTAAAA CTGCGGGTTG GTTGGCAGAA
451 CTTTCCACA AGAATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA
501 ATATAAACGC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG
551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTCCTT ATGACACATC
651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAAACTGG
701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG
751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTGAGCT GGCAATTTAA
801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC
851 AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGTCGGCAC AGCAATTGGG
901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC
951 CCGCCGCGCA TTGAAAAAGC CCCGAATTTT CCAATCAAGG AAATGGGCAA
1001 GCGGTTTTCA GGTAGGCTAT ACGTTTTAA

This corresponds to the amino acid sequence [<SEQ ID 606; ORF142-1>] (SEQ ID NO: 606; ORF142-1):

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH
51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAVSGLSE VYDYNKSYN
101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE
151 LSHKEYIGRS TADFKLYKR GTGMKDALRA PEEAFGEGTS RMKIWTASAD
201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
251 SAERGWIWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWSL GQTLVGTAIG
301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 (SEQ ID NO: 604) shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) (SEQ ID NO: 608) from *N.gonorrhoeae*:

-444-

```

orf142.pep                                     QSAKWLSGQTLVGTAIGIRGQIKLGGNLHY   30
      |||||||||||||:|||||||||||||||
orf142ng      RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY   313

orf142.pep      DIFTGRALKKPEFFQSRKWASGFQVGYTF      59
5      |||||||||||||:||||:||||:||||:|
orf142ng      DIFTGRALKKPEYFQTKKWVTGFQVGYSF      342

```

The complete length ORF142ng nucleotide sequence [<SEQ ID 607>] (SEQ ID NO: 607) is:

```

10      1  ATGGATAATT  CGGGTAGTGA  GGCGACAGGA  AAATACCAAG  GAAATATCAC
      51  TTTCTCTGCC  GACAATCCTT  TTGGACTGAG  TGATATGTTC  TATGTAAATT
      101  ATGGACGTTC  AATTGGCGGT  ACGCCCGATG  AGGAAAATTT  TGACGGCCAT
      151  CGCAAAGAAG  GCGGATCAAA  CAATTACGCC  GTACATTATT  CAGCCCCTTT
      201  CGGTAAATGG  ACATGGGCAT  TCAATCACAA  TGGCTACCGT  TACCATCAGG
      251  CGGTTTCCGG  ATTATCGGAA  GTCTATGACT  ATAATGGAAA  AAGTTACAAC
      15      301  ACTGATTTCG  GCTTCAACCG  CCTGTTGTAT  CGTGATGCCA  AACGCAAAAC
      351  CTATCTCAGT  GTAAAACGTG  GGACGAGGGA  AACAAAAAGT  TACATTGATG
      401  ATGCCGAACT  GACTGTACAA  CGGCGTAAAA  CCACAGGTTG  GTTGGCAGAA
      451  CTTTCCCACA  AAGGATATAT  CGGTCGCAGT  ACGGCAGATT  TTAAGTTGAA
      501  ATATAAACAC  GGCACCGGCA  TGAAAGATGC  TCTGCGCGCG  CCTGAAGAAG
      20      551  CCTTTGCGCA  AGGCACGTCA  CGTATGAAAA  TTTGGACGGC  ATCGGCTGAT
      601  GTAAATACTC  CTTTTCAAA  CGGTAAACAG  CTATTTGCCT  ATGACACATC
      651  CGTTCATGCA  CAATGGAACA  AAACCCCGCT  AACATCGCAA  GACAACTGG
      701  CTATCGGCGG  ACACCACACC  GTACGTGGCT  TCGACGGTGA  AATGAGTTTG
      751  CCTGCCGAGC  GGGGATGGTA  TTGGCGCAAC  GATTGAGCT  GGCAATTTAA
      25      801  ACCAGGCCAT  CAGCTTTATC  TTGGGGCTGA  TGTAGGACAT  GTTTCAGGAC
      851  AATCCGCCAA  ATGGTTATCG  GGCCAACTC  TAGCCGGCAC  AGCAATTGGG
      901  ATACGCGGGC  AGATAAAGCT  TGGCGGCAAC  CTGCATTACG  ATATATTTAC
      951  CGGCCGTGCA  TTGAAAAAGC  CCGAATATTT  TCAGACGAAG  AAATGGGTAA
      1001  CGGGGTTTCA  GGTGGGTTAT  TCGTTTGA
30

```

This encodes a protein having amino acid sequence [<SEQ ID 608>] (SEQ ID NO: 608):

```

1  MDNSGSEATG  KYQGNITFSA  DNPFGLSDMF  YVNYGRSIGG  TPDEENFDGH
35      51  RKEGGSNNYA  VHYSAPFGKW  TWAFNHNGYR  YHQA VSGLSE  VDYNGKSYN
      101  TDFGNRLLY  RDAKRKTYLS  VKLWTRETKS  YIDDAELTVQ  RRKTGWLAE
      151  LSHKGYIGRS  TADFKLYKH  GTGMKDALRA  PEEAFGEGTS  RMKIWTASAD
      201  VNTPFQIGKQ  LFAYDTSVHA  QWNKTPLTSQ  DKLAIGGHHT  VRGFDGEMSL
      251  PAERGWYWRN  DLSWQFKPGH  QLYLGADVGH  VSGQSAKWLS  GQTLAGTAIG
      301  IRGQIKLGGN  LHYDIFTGRA  LKKPEYFQTK  KWTGFGQVGY  SF*

```

40 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng (SEQ ID NO: 608) and ORF142-1 (SEQ ID NO: 606) show 95.6% identity over 342aa overlap:

```

45      orf142-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
      |||||||||||||:|||||||||||||||
      orf142ng-1   MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA

      orf142-1.pep VHYSAPFGKWTWAFNHNGYRYHQA VSGLSEVVDYNGKSYNTDFGNRLLYRDAKRKTYLG
      |||||||||||||:|||||||||||||||

```

orf142ng-1	VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKGSYNTDFGFNRLLYRDAKRKYLS
orf142-1.pep	VKLWMRETKSYIDDAELTVQRRKTAGWLAE LSHKEYIGRSTADFKLKYKRGTMKDALRA : :
orf142ng-1	VKLWTRETKS YIDDAELTVQRRKTTGWLAE LSHKGYIGRSTADFKLKYKHGTGMKDALRA
orf142-1.pep	PEEAFGEGETSRMKIWTASADVNTPFQIGQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
orf142ng-1	PEEAFGEGETSRMKIWTASADVNTPFQIGQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
orf142-1.pep	VRGFDGEMSLSAERG WYWRNDLSWQFKPGHQLYLGADVGVHSVGSQAKWLSGQT LVGTAIG :
orf142ng-1	VRGFDGEMSLPAERGWYWRNDLSWQFKPGHQLYLGADVGVHSVGSQAKWLSGQTLAGTAIG
orf142-1.pep	IRGQIKLGGNLHYDI FTGRALKKP EFFQS R KWAS GFQVG YTF : :
orf142nq-1	IRGQIKLGGNLHYDI FTGRALKKPEY F QTKKWVTG FQVGYSF

- 15 In addition, ORF142ng (SEQ ID NO: 608) is homologous to the HecB protein (SEQ ID NO: 1149)
of *E.chrysanthemi*:

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
Score = 119 bits (295), Expect = 3e-26
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)

20 Query: 2 DNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
 DNSG ++TG+ Q N + + DN FGL+D ++++ G S + + D + G \
 Sbjct: 230 DNSGQKSTGEEQLNGSLALDNVFLADQWFISAGHS---SRFATSHDAESLQAG----- 280

25 Query: 62 HYSAPFGKWTWAFNHNHGYRYHQAVSGLSEVYDYNKGKSYNTDFGNRLLYRDAKRKTYLSV 121
+S P+G W +N++ RY + G S F +R+++RD KT ++
Sbjct: 281 -FSMPYGYWNLGYNYSQSRYNRTFINRDFPWHSTGDS DTHRFSLSRVVFRDGMTKTAIAG 339

Query: 122 KLWTRKTSYIDDAELTVQRRKTTGWLAE LSHKGYIGRSTADFKLYKHGTGMKDALRAP 181
R +Y++ + L RK + ++H + A F Y G +
Sbjct: 340 TFSQRTGNNYLNGSLLPSSSRKLSSVSLGVNHSQKLWGGLATFNPTYNRGVRWLGSSETDT 399

30 Query: 182 EEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241
+++ E + WT SA P Y S++ Q++ L ++L +GG ++
Sbjct: 400 DKSADPEPRAEFNKWTLASYYHPV---TDSITYLGLSYGQYSARALYGSEQLTLGGESSI 456

Query: 242 RGF D G E M S L P A E R G W Y W R N D L S W Q F K P - - - - G H Q L Y L G A - D V G H V S G Q S A K W L S G Q T L A G 296
 R G F E R G Y W R N + L + W Q G + + + A D G H + + + L G
 Sbjct: 457 R G F - R E Q Y T S G N R G A Y W R N E L N W Q A W Q L P V L G N V T F M A A V D G G H L Y N H K Q D N S T A A S L W G 515

35 Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWWTGFQVGYSF 342
 A+G+ + L + G + P + Q V G++VG SF
 Sbjct: 516 GAVGMTVASRW--LSQQVTVGWPIISYPAWLQPDMTMVVGYRVGLSF 558

- On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 73

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 609>] (SEQ ID NO: 609):

```

5      1  ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GgCGGACACC
      51  GCCGACATCG ATACCGCTTT GAACCTGTTG TACCGTTTGC AAAA5ACTCGA
     101  ATTCCTCTAT GGCGATGAAA ACGGTCATTC AGACGGCATC AATTGwCGG
     151  ACGAGCAATT GCCGTTGCTG ATGGAACAAT TGTCCGGCAG CGGTAAGGCG
     201  TTATTGGTCG ATCGGAACGG TCTGTATCTT GCCAACGCCA ATTTCCATCA
     251  TGAGGCGGCG GAAGAGTTGG GGTGTGTGGC GGCAGAAGTC GCACAGATGG
    10   301  AAAAGAAATA CCGGCTGCTG ATTAAGAACA AC. .

```

This corresponds to the amino acid sequence [<SEQ ID 143>] (SEQ ID NO: 610; ORF143):

```

15      1  MRTKWSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
      51  EQLPLLMEQL SGSGKALLVD RNgLYLANAN FHHEAAEELG LLAAEVAQME
     101  KKYRLLIKNN ..

```

Further work revealed the complete nucleotide sequence [<SEQ ID 611>] (SEQ ID NO: 611):

```

20      1  ATGGAATCAA CACTTTCACT ACAAGCAAAT TTATATCCCC GCCTGACTCC
      51  TGCCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCCGTAAAA
     101  CTTTGTGTGCA CAGCCTGTTG AAAGCAGATG CGGACGAAAT GGTCA5GAGT
     151  GAGAAGCTGC TTA5CTTGGGC GGACACCGCC GACATCGATA CCGCTTTGAA
     201  CCTGTTGTAC CGTTTGC5AAA AACTCGAATT CCTCTATGGC GATGAAAACG
     251  GTCATT5CAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
     301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
     351  GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT
     401  TGTTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCTGATT
     451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
     501  CGGTCAGAGC GAATTGACAT TTTTCCCAT5T GTATATCGGT TCAACCAAAT
     551  TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTGTGT
    30   601  ACTTTGGTAA GGATTTTATA CCGCCGT5TAC AGCAACCGCG TGTA5A

```

This corresponds to the amino acid sequence [<SEQ ID 612; ORF143-1>] (SEQ ID NO: 612; ORF143-1):

```

35      1  MESTLSLQAN LYPR5LTPAGA FYAVSSDAPS AGKTL5LHSL5L KADADEMVSS
      51  EKLLTWADTA DIDTALNLLY RLQKLEFLYG DENGHS5DGIN LSDEQLPLLM
     101  EQLSGSGKAL LVDRNGLYLA NANF5HHEAAE ELGL5LAAEVA QMEKKYRLLI
     151  KNNLYINNNA WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV
     201  TLVRILYRRY SNRV*

```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF143 (SEQ ID NO: 610) shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) (SEQ ID NO: 614) from strain A of *N. meningitidis*:

```

5      orf143.pep      10      20      30
                        MRTKWSAVRSC TWADTADIDTALNLLYRLQKLEFL
                        | : : | | | | | | | | | | | | | | | |
orf143a      GAFYAVSSDXPSAGKTL LHSLLKADADEMVSSSEKLLTWAXTADIDTALNLLYRLQKLEFL
                20      30      40      50      60      70

10     orf143.pep      40      50      60      70      80      90
                        YGDENGHS DGINLXDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHEAAEELGLLAAE
                        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf143a      YGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHEAAEELGLLAAE
                80      90      100     110     120     130

15     orf143.pep      100     110
                        VAQMEKKYRL LIKNN
                        | | | | | | | | | |
orf143a      VAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFPFLYIGSTKFILVIGGIPDLGKEA
                140     150     160     170     180     190

```

The complete length ORF143a nucleotide sequence [<SEQ ID 613>] (SEQ ID NO: 613) is:

```

20      1  ATGGAATCAA CANTTTCACT ACAAGCAAAT TTATATCNCC GCCTGACTCC
        51  TGCCCGTGCA TTTTATGCCG TATCCAGCGA TGNCCCCAGT GCCCGTAAAA
       101  CTTTGTGTGA CAGCCTGTTG AAAGCGGATG CGGACGAAAT GGTNAGCAGT
       151  GAGAAGCTGC TTACCTGGGC GGANACCGCC GACATCGATA CCGCTTTGAA
       201  CCTGTTGTAC CGTTTGCAA AACTCGAATT CCTCTATGGC GATGAAAACG
       25      251  GTCATTGAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
        301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
       351  GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT
       401  TGTTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCNNATT
       451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
       501  CGGTCAGAGC GAATTGACAT TTTTCCCAT TATATCGGT TCAACCAAAT
       551  TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTGTT
       601  ACTTTGGTAA GGATNTTATA CCNCCNGTTA CAGCAACCGC GTGTAAAACT
       651  TGGGAGAGAG GANGGGTTAT GCAGCAATTA TTGA

```

35 This encodes a protein having amino acid sequence [<SEQ ID 614>] (SEQ ID NO: 614):

```

40      1  MESTXSLQAN LYXRLTPAGA FYAVSSDXPS AGKTL LHSLL KADADEMVSS
        51  EKLLTWAXTA DIDTALNLLY RLQKLEFLYG DENGHS DGIN LSDEQLPLL
       101  EQLSGSGKAL LVDRNGLYLA NANFHEAAE ELGLLAAEVA QMEKKYRLXI
       151  KNNLYINNNA WGVCDPSGQS ELTFPFLYIG STKFILVIGG IPDLGKEAFV
       201  TLVRXLYXXL QQPRVKLGRE XGLCSNY*

```

ORF143a (SEQ ID NO: 614) and ORF143-1 (SEQ ID NO: 612) show 97.1% identity in 207 aa overlap:

```

45     orf143a.pep      MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTL LHSLLKADADEMVSSSEKLLTWAXTA
                        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf143-1      MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTL LHSLLKADADEMVSSSEKLLTWADTA

orf143a.pep      DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA

```

```

      |||
orf143-1  DIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
      |||
orf143a.pep  NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFPLYIG
5  orf143-1  NANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPSGQSELTFFPLYIG
      |||
orf143a.pep  STKFILVIGGIPDLGKEAFVTLVRXLY
      |||
orf143-1  STKFILVIGGIPDLGKEAFVTLVRILY

```

Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF143 (SEQ ID NO: 610) shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) (SEQ ID NO: 616) from *N.gonorrhoeae*:

```

orf143.pep  MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYGDENGHSDGINLXDEQLPLLMEQL  60
      |||
orf143ng    MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQL  60
15  orf143.pep  SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIKNN  110
      |||
orf143ng    SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLIRNNLYINNNAWGV  120

```

20 An ORF143ng nucleotide sequence [<SEQ ID 615>] (SEQ ID NO: 615) was predicted to encode a protein having amino acid sequence [<SEQ ID 616>] (SEQ ID NO: 616):

```

      1  MRTKWSAVRS  CSRADTADID  TALNLLYRLQ  KLEFLYGDEN  GHSDGINLSD
      51  EQLPPLLMEQL  SGSGKALLVD  RNgLYLANAN  FHHEAAEELG  LLAAEVAQME
     101  KKYRLLIRNN  LYINNNAWGV  CDPSGQSELT  FFPLYIGSTK  FILVIAGIPD
     151  LSKGGICYFG  KDFIPPLQQP  RVKLTGGGIM  RQLLISILED  LNNTSTDIIA
     201  SAVISTDGLP  MATMLPSHLN  SDRVGAISAT  LLALGSRVQ  ELACGELEQV
     251  MIKGKSGYIL  LSQAGKDAVL  VLVAKETGRL  GLILLDAKRA  ARHIAEAI*

```

Further work revealed the following gonococcal DNA sequence [<SEQ ID 617>] (SEQ ID NO: 617):

```

      1  ATGGAATCAA  CACTTTCACT  ACAAGCGAAT  TTATATCCCT  GCCTGACTCC
     30  TGCCGGTGCA  TTTTATGCCG  TATCCAGCGA  TGCCCCCAGT  GCCGGTAAAA
     101  CTTTGTGTCG  CAGCCTGTTG  AAAGCGGATG  CGGACGAAGT  GGTCAGCAGT
     151  GAGAAGCTGC  TCGCGGCGGA  CACCGCCGAC  ATCGATACCG  CTTTGAACCT
     201  GTTGTACCGT  TTGCAAAAAC  TCGAATTCCT  CTATGGCGAT  GAAAACGGTC
     251  ATTCAGACGG  CATCAATTG  TCGGACGAGC  AATTGCCGTT  GCTGATGGAA
     301  CAATTGTCCG  GCAGCGGTAA  GGCATTATTG  GTCGATCGGA  ACGGTCTGTA
     351  TCTTGCCAAC  GCCAATTTC  ATCATGAGTC  GGCGGAAGAG  TTGGGGTTGT
     401  TGGCGGCAGA  AGTCGCACAG  ATGGAAAAGA  AATACCGGCT  GCTGATTAGG
     451  AACAACTGT  ATATCAACAA  TAACGCTTGG  GGCGTTTGCG  ATCCTTCCGG
     501  TCAGAGCGAA  TTGACATTT  TCCCATTTGA  TATCGGTTCA  ACCAAATTTA
     551  TTTTGGTTAT  CGCCGGCATT  CCCGATTGTA  GCAAAGAGGC  ATTTGTTACT
     601  TTGGTAAGGA  TTTTATACCG  CCGTTACAGC  AACCGCGTGT  AA

```


This corresponds to the amino acid sequence [<SEQ ID 618; ORF143ng-1>] (SEQ ID NO: 618; ORF143ng-1):

```

5      1  MESTLSLQAN LYPCLTPAGA FYAVSSDAPS AGKTLRLSLL KADADEVVSS
      51  EKLLAADTAD IDTALNLLYR LQKLEFLYGD ENGHSDGINL SDEQLPLLME
      101  QLSGSGKALL VDRNGLYLAN ANFHHEAEE LGLLAAEVAQ MEKKYRLLR
      151  NNLYINNNAW GVCDPSGQSE LTFFFLYIGS TKFILVIAGI PDLSKEAFVT
      201  LVRILYRRYS NRV*

```

ORF143ng-1 (SEQ ID NO: 618) and ORF143-1 (SEQ ID NO: 612) show 95.8% identity in 214 aa overlap:

```

      orf143ng-1.pep MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSEKLLA-ADTA 59
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf143-1      MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSEKLLTWADTA 60

      orf143ng-1.pep DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 119
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf143-1      DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 120

      orf143ng-1.pep NANFHHEAEEELGLLAAEVAQMEKKYRLLRNNLYINNNAWGVCDPSGQSELTFFFLYIG 179
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf143-1      NANFHHEAEEELGLLAAEVAQMEKKYRLLRNNLYINNNAWGVCDPSGQSELTFFFLYIG 180

      orf143ng-1.pep STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV 213
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf143-1      STKFILVIGGIPDLGKEAFVTLVRILYRRYSNRV 214

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 619>] (SEQ ID NO: 619):

```

30      1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
      51  GTTTCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGT
      101  CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCGCGTGTGC
      151  ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCGGTGTTTC ACCGTGTGTC
      201  GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG
35      251  ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
      301  ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCG
      351  GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAA wTyCCAGCGT
      401  CCGTGGATG..

```

This corresponds to the amino acid sequence [<SEQ ID 620; ORF144>] (SEQ ID NO: 620; ORF144):

```

1  MTFLQRLQGL ADNKICAFW FVRRFDEER VPQXAASMTF TLLALVPVL
51 TVMVAVASIF PVFDRWDSF VSFVNQTIVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM...
```

Further work revealed the complete nucleotide sequence [<SEQ ID 621>] (SEQ ID NO: 621):

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51 GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTTC ACCGTGGTTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
15 351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGCTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTATGACG CTTTGTCTGT GGGGGCTGTA CCGCTTCGTG
20 601 CCAAACCGCT TCGTTCGCGC GCGGCAGGCG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTTGTG CTGGAACCGC CGCGCTCCCT CTTCACTTGG TATATGGGCA
701 ATTTGACGCG CTACCGCTCG ATTTACGCGC CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
25 851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAATACTCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTTCTG AGTTTCTGACG
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACCG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG
30 1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACGGCAGTA G
```

This corresponds to the amino acid sequence [<SEQ ID 622; ORF144-1>] (SEQ ID NO: 622; ORF144-1):

```

1  MTFLQRLQGL ADNKICAFW FVRRFDEER VPQXAASMTF TLLALVPVL
51 TVMVAVASIF PVFDRWDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFMT LLLWGLYRFV
40 201 PNRFPVPAQA FVGALATAFC LETARSLFTW YMGNFDDYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLAHGYIY SGRQGWVLKT
351 GADSIELNEL FKLFFVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
401 QAKKRQ*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 (SEQ ID NO: 620) shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) (SEQ ID NO: 624) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
5	orf144.pep	MTFLQRLQGLADNKICAFWVRRFDEERVQPXAASMTFTTLLALVPVLTVMVAVASIF					
	orf144a	MTFLQRLQGLADNKICAFWVRRFDEERVQAAAASMTFTTLLALVPVLTVMVAVASIF					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf144.pep	PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID					
	orf144a	PVFDRWSDSFVSFVNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID					
		70	80	90	100	110	120
		130					
15	orf144.pep	NTFNRIWRVXXQRPWM					
	orf144a	NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL					
		130	140	150	160	170	180

The complete length ORF144a nucleotide sequence [<SEQ ID 623>] (SEQ ID NO: 623) is:

20	1	ATGACCTTTT	TACAACGTTT	GCAAGGTTTG	GCAGACAATA	AAATCTGTGC
	51	GTTTGCGATGG	TTCGTCGTCC	GCCGCTTTGA	TGAAGAACGC	GTACCGCAGG
	101	CGGCGGCAAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTGCTG
	151	ACCGTGATGG	TGGCGGTGCG	TTCGATTTTC	CCCGTGTTTCG	ACCGNTGGTC
	201	GGATTCTGTT	GTCTCTTTCG	TCAACCAAAC	CATTGTGCCG	CAGGGCGCGG
25	251	ACATGGTNTT	CGACTATATC	AATGCGTTCC	GCGAGCAGGC	GAACCGGCTG
	301	ACGGCAATCG	GCAGCGTGAT	GCTGGTCGTT	ACCTCGCNGA	TGCTGATTCG
	351	GACGATAGAC	AATACGTTC	ACCGCATCTG	GCGGTCAAT	TCCAGCGTC
	401	CGTGGATGAT	GCAGTTTCTC	GTCTATTGGG	CTTTACTGAC	GTTCCGGCCG
	451	CTGTCTTTGG	GCGTGGGCAT	TTCTTTATN	GTCCGCTCGG	TACAGGATGC
30	501	CGCGCTTGCC	TCAGGTGCGC	CGCAGTGGTC	GGGCGCGTTG	CGAACGGCGG
	551	CGACGCTGAN	CTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTNCGTG
	601	CCAAACCGCT	TCGTTCCCGC	GCGGCANGCG	TTTGTCGGGG	CTTTGGCAAC
	651	AGCGTCTCTG	CTGGAAACCG	CGCGTTCCTT	CTTTACTTGG	TATATGGGCA
	701	ATTTTCGACG	CTACCGCTCG	ATTTACGGNG	CGTTTGCCGC	CGTGCCGTTT
35	751	TTTCTGTTGT	GGCTGAACCT	GTTGTGGACG	CTGGTCTTGG	GCGGCGCGGT
	801	GCTGACTTCT	TACTCTCCT	ACTGGCAGGG	AGAAGCGTTC	CGCAGGGNCT
	851	TCGACTCGCG	CGGACGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
	901	GATGCGGCGC	AAAAAGAAGG	CNAAGCCTTG	CCTGTTTCAGG	AGTTCAGACG
	951	GCATATCAAT	ATGGGCTACG	ACGAGTTGGG	CGAGCTTTTG	GAAAAGCTGG
40	1001	CGCGGCACGG	CTACATCTAT	TCCGGCAGAC	AGGGTTGGGT	GTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAACGAACTC	TTCAAGCTCT	TCGTTTACCG
	1101	TCCGTTGCCT	GTGGAAAGGG	ATCATGTGAA	CCAAGCTGTC	GATGCGGTAA
	1151	TGATGCCGTG	TTTGACACT	TTGAACATGA	CGCTGGCAGA	GTTTGACGCT
45	1201	CAGGCGAAAA	AACAGCAGCA	ATCTTGA		

This encodes a protein having amino acid sequence [<SEQ ID 624>] (SEQ ID NO: 624):

	1	MTFLQRLQGL	ADNKICAFW	FVRRFDEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSVFNQITVP	QGADMVFDYI	NAFREQANRL
50	101	TAIGSVMLVV	TSXMLIRTID	NTFNRIWRVN	SQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFX	VGSVQDAALA	SGAPQWSGAL	RTAATLXFMT	LLWGLYRXV
	201	PNRFVPARXA	FVGALATAFC	LETARSLFTW	YMGNFDGYRS	IYGAFAAVPF

251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRXFDSRGRF DDVLKILLLL
 301 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYYIY SGRQGWVLKT
 351 GADSIELNEL FKL FVYRPLP VERDHVNQAV DAVMMPCLQT LNMTLAEFDA
 401 QAKKQQQS*

5

ORF144a (SEQ ID NO: 624) and ORF144-1 (SEQ ID NO: 622) show 97.8% identity in 406 aa overlap:

10	orf144a.pep	MTFLQRLQGLADNKICAFAWFVRRFDEERVQAAAASMTFTTLLALVPVLTVMVAVASIF	
	orf144-1	MTFLQRLQGLADNKICAFAWFVRRFDEERVQAAAASMTFTTLLALVPVLTVMVAVASIF	
	orf144a.pep	PVFDRWSDSFVSFVNQTIVPQADMVFDYINAFREQANRLTAIGSVMLVVT SXMLIRTID	
	orf144-1	PVFDRWSDSFVSFVNQTIVPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	
15	orf144a.pep	NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL	
	orf144-1	NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL	
	orf144a.pep	RTAATLXFMTLLWGLYRXVPNRFVPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS	
	orf144-1	RTAATLTFMTLLWGLYRFVPNRFVPARQAFVGALATAFCLETARSLFTWYMGNFDGYRS	
20	orf144a.pep	IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL	
	orf144-1	IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL	
25	orf144a.pep	DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIELNEL	
	orf144-1	DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIELNEL	
	orf144a.pep	FKLFVYRPLPVERDHVNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS 408	
	orf144-1	FKLFVYRPLPVERDHVNQAVDAVMTPCQLTLNMTLAEFDAQAKKRQ 406	

Homology with a predicted ORF from *N.gonorrhoeae*

30 ORF144 (SEQ ID NO: 620) shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) (SEQ ID NO: 626) from *N.gonorrhoeae*:

35	orf144 .pep	MTFLQRLQGLADNKICAFAWFVRRFDEERVQXAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFAWFVIRRFSEERVQAAAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144 .pep	PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQTIVPQADMVFDYIDAFRQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144 .pep	NTFNRIWRVXXQRPWM	136
40	orf144ng	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence [<SEQ ID 625>] (SEQ ID NO: 625) is predicted to encode a protein having amino acid sequence [<SEQ ID 626>] (SEQ ID NO: 626):

```

1  MTFLQCWQGS  ADNKICAFW  FVIRRFSEER  VPQAAASMTF  TLLALVPVL
5  51  TVMVAVASIF  PVFDRWDSF  VSFVNQTIVP  QGADMVFDYI  DAFRDQANRL
101 TAIGSVMLVV  TSLMLIRTID  NAFNRIWRVN  TQRPWMMQFL  VYWALLTFGP
151 LSLGVGISFM  VGSVQDSVLS  SGAQQWADAL  KTAARLAFMT  LLLWGLYRFV
201 PNRFVPARQA  FVGALITAF  LETARFLFTW  YMGNFDDGYRS  IYGAFVAVPF
251 FLLWLNLLWT  LVLGGAVLTS  SLSYWQGEAF  RRGFDSRGRF  DDVLKILLLL
10 301 DAAQKEGRTL  SVQEFRRHIN  MGYDELGELL  EKLARYGYIY  SGRQGWVLKT
351 GADSIELSEL  FKLTVYRPLP  VERDHVNQAV  DAVMTPCLQT  LNMTLAEFDA
401 QAKKQQQS*

```

Further work revealed the following gonococcal DNA sequence [<SEQ ID 627>] (SEQ ID NO: 627):

```

15 1  ATGACCTTTT  TACAACGTTG  GCAAGGTTTG  GCGGACAATA  AAATCTGTGC
51  ATTTGCATGG  TTCGTCATCC  GCCGTTTCAG  TGAAGAGCGC  GTACCGCAGG
101 CAGCGGCGAG  CATGACGTTT  ACGACACTGC  TGGCACTCGT  CCCCGTACTG
151 ACCGTAATGG  TCGCGGTCGC  TTCGATTTTC  CCCGTGTTTC  ACCGCTGGTC
201 GGATTCGTTT  GTCTCCTTCG  TCAACCAAAC  CATTGTGCGC  CAGGGCGCGG
20 251 ATATGGTGTT  CGACTATATC  GACGCATTCC  GCGATCAGGC  AAACCGGCTG
301 ACCGCCATCG  GCAGCGTGAT  GCTGGTCGTA  ACCTCGCTGA  TGCTGATTCTG
351 GACGATAGAC  AATGCGTTCA  ACCGCATCTG  GCGGGTTAAC  ACGCAACGCC
401 CCTGGATGAT  GCAGTTCCTC  GTTTATTGGG  CGTTGCTGAC  TTTCGGGCCT
451 TTGTCTTTGG  GTGTGGGCAT  TTCCTTTATG  GTCGGGTCGG  TTCAAGACTC
25 501 CGTACTCTCC  TCCGGAGCGC  AACAATGGGC  GGACGCGTTG  AAGACGGCGG
551 CAAGGCTGGC  TTTCATGACG  CTTTGTGCTG  GGGGGCTGTA  CCGCTTCGTG
601 CCCAACCGCT  TCGTGCCCGC  CCGGCAGGCG  TTTGTGCGAG  CTTTGATTAC
651 GGCATTCTGC  CTGGAGACGG  CACGTTTCCT  GTTCACCTGG  TATATGGGCA
30 701 ATTTGACGCG  CTACCGCTCG  ATTTACGCGC  CATTGTCCGC  CGTGCCGTTT
751 TTCTGCTGT  GTTAAACCT  GCTGTGGACG  CTGGTCTTGG  GCGGGCGGCT
801 GCTGACTTCG  TCGCTGTCTT  ATTGGCAGGG  CGAGGCCTTC  CGCAGGGGAT
851 TCGACTCGCG  CGGACGGTTT  GACGACGTGT  TGAAAATCCT  GCTGCTTCGT
901 GATGCGGCGC  AAAAAGAAGG  CCGAACCTTG  TCCGTTACAG  AGTTCAGACG
951 GCATATCAAT  ATGGGTACG  ATGAATTGGG  CGAGCTTTTG  GAAAAGCTGG
35 1001 CGCGGTACGG  CTATATCTAT  TCCGGCAGAC  AGGGCTGGGT  TTTGAAAACG
1051 GGGGCGGATT  CGATTGAGTT  GAGCGAATC  TTCAAGCTCT  TCGTGTACCG
1101 CCCGTTGCct  gtggaAAGGG  ATCATGTGAA  CCAAGCTGtc  gaTGCGGTAA
1151 TGAcgccgtG  TTTGCAGACT  TTGAACATGA  CGCTGGCGGA  GTTTGACGCT
40 1201 CAGgcgAAAA  AACAGCAGCA  GTCTTGA

```

This encodes a variant of ORF144ng, having the amino acid sequence [<SEQ ID 628; ORF144ng-1>] (SEQ ID NO: 628; ORF144ng-1):

```

45 1  MTFLQRWQGL  ADNKICAFW  FVIRRFSEER  VPQAAASMTF  TLLALVPVL
51  TVMVAVASIF  PVFDRWDSF  VSFVNQTIVP  QGADMVFDYI  DAFRDQANRL
101 TAIGSVMLVV  TSLMLIRTID  NAFNRIWRVN  TQRPWMMQFL  VYWALLTFGP
151 LSLGVGISFM  VGSVQDSVLS  SGAQQWADAL  KTAARLAFMT  LLLWGLYRFV
201 PNRFVPARQA  FVGALITAF  LETARFLFTW  YMGNFDDGYRS  IYGAFVAVPF
251 FLLWLNLLWT  LVLGGAVLTS  SLSYWQGEAF  RRGFDSRGRF  DDVLKILLLL
50 301 DAAQKEGRTL  SVQEFRRHIN  MGYDELGELL  EKLARYGYIY  SGRQGWVLKT
351 GADSIELSEL  FKLTVYRPLP  VERDHVNQAV  DAVMTPCLQT  LNMTLAEFDA
401 QAKKQQQS*

```

5 orf144ng-1.pep MTFLLQRWQGLADNKKICAFWFVIRRFSEERVPPQAAAASMTFTTLLALVPVLTVMVAVASIF
 orf144-1 MTFLLQRWQGLADNKKICAFWFVIRRFSEERVPPQAAAASMTFTTLLALVPVLTVMVAVASIF

orf144ng-1.pep PVFDRWSDSFVSFVNQTIIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
 orf144-1 PVFDRWSDSFVSFVNQTIIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID

10 orf144ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
 orf144-1 NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL

orf144ng-1.pep KTAARLAFMTLLWGLYRFVFNRFVPPARQAFVGGALITAFCLLETARFLFTWYMGNFDDGYRS
 orf144-1 RTAATLTFMTLLWGLYRFVFNRFVPPARQAFVGGALITAFCLLETARSLFTWYMGNFDDGYRS

15 orf144ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
 orf144-1 IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL

orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELEKLARYGYIYSGRQGWVLKTGADSIELSEL
 orf144-1 DAAQKEGKALPVQEFRRHINMGYDELGELEKLARHGYIYSGRQGWVLKTGADSIELNEL

20 orf144ng-1.pep FKLFFVYRPLPVERDHVNQAVDAVMTPCLOTNMTLAEFDAQAKKQQQS
 orf144-1 FKLFFVYRPLPVERDHVNQAVDAVMTPCLOTNMTLAEFDAQAKKRQ

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 75

The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 629](#)] ([SEQ ID NO: 629](#)):

35

1	..AGACACGCCC	GCCGCATCCG	CATCGACACC	GCCATCAACC	CCGAACTGGA
51	AGCCCTCGCC	GAACACCTCC	ACTACCAATG	GCAGGGCTTC	CTCTGGCTCA
101	GCACCGATAT	GCGTCAGGAA	ATTTCCGCCC	TCGTCA TCCT	GCTGCAACGC
151	ACCTGCCCGCA	AATGGCTGGA	TGCCCAACAA	CGCCAACACC	TGCGCCAAAG
201	CCTGCCTGAA	ACACGGGAAC	ACGGCTGA		

This corresponds to the amino acid sequence [<SEQ ID 630; ORF146>] (SEQ ID NO: 630; ORF146):

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51  TRRKWLDAHE RQHLRQSLLE TREHG*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 631>] (SEQ ID NO: 631):

```

1  ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
51  CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTCGGCGGGG
101 CCGTCTGTGT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
151 GAGTGGATAG GGATGACCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTTCA
201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
401 CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
451 CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
601 AGGCGCATGA CCCGCGAACG CCTCGAGGAG AACATGGCGA AAATGCGCCA
651 AATCAACGCA CGCATGGTCA AAAGCCGCAG CCATCTCGCC GCCACATCGG
701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCCAC
751 CGTAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
851 TCACACTGCT CCAAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAG
1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence [<SEQ ID 632; ORF146-1>] (SEQ ID NO: 632; ORF146-1):

```

1  MNTSQRNRLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
51  EWIGMTVFVV LGMLQFQAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLD SG
151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
251 RKIVNTTELL LTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
351 TRRKWLDAHE RQHLRQSLLE TREHG*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF146 (SEQ ID NO: 630) shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) (SEQ ID NO: 634) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF146a nucleotide sequence [<SEQ ID 633>] (SEQ ID NO: 633) is:

	1	ATGAACACCT	CGCAACGCAA	CCGCTCGTGC	AGCCGCTGGC	TCAACTCCTA
15	51	CGAACGCTAC	CGCTACCGCC	GCCTCATCCA	CGCCGTCGGG	CTCGGCGGGG
	101	CCGTCCTGTT	CGCCACCGCC	TCCGCCCGGC	TGCTCCACCT	CCAACACGGC
	151	GAGTGGATAG	GGATGACCGT	CTTCGTCGTC	CTCGGCATGC	TCCAGTTTCA
	201	AGGGGCGGATT	TACTCCAAGG	CGGTGGAACG	TATGCTCGGC	ACGGTCAATC
20	251	GGCTGGGCGC	GGGTTTGGGC	GTTTTATGGC	TGAACACGCA	TTATTTCCAC
	301	GGCAACCTCC	TCTTCTACCT	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG
	351	CTGGGCGGCG	GTCGGCAAAA	ACGGCTACGT	CCCTATGCTG	GCGGGGCTGA
	401	CGATGTGCAT	GCTCATCGGC	GACAACGGCA	GCGAATGGTT	CGACAGCGGC
25	451	CTGATGCGCG	CGATGAACGT	CCTCATCGGC	GCGGCCATCG	CCATCGCGCT
	501	CGCCAAACTG	CTGCCGCTGA	AATCCACACT	GATGTGGCGT	TTTACTTCTT
	551	CCGACAACTC	GACCGACTGC	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC
	601	AGGCGCATGA	CCCGCGAACG	CCTCGAAGAG	AACATGGCGA	AAATGCGCCA
30	651	AATCAACGCA	CGCATGGTCA	AAAGCCCGAG	CCACTCGGCC	GCCACATCGG
	701	GCGAAAGCCG	CATCAGCCCC	GCCATGATGC	AAGCCATGCA	GCACGCCCCA
	751	CGTAAATTG	TCAACACCAC	CGAGCTGTCT	CTGACCACCG	CCGCCAAGCT
	801	GCAATCTCCC	AAACTCAACG	GCAAGCGAAAT	CCGGCTGCTT	GACCGCCACT
35	851	TCACACTGCT	CCAAACCGAC	CTGCAACAAA	CCGTCGCCCT	TATCAACGGC
	901	AGACACGCCC	GCCGCATCCG	CATCGACACC	GCCATCAACC	CCGAACTGGA
	951	AGCCCTCGCC	GAACACCTCC	ACTACCAATG	GCAGGGCTTC	CTCTGGCTCA
	1001	GACCAATAT	GCGTCAGGAA	ATTTCCGCCC	TCGTATCCTT	GCTGCAACGC
	1051	ACCCGCCGCA	AATGGCTGGA	TGCCACGAA	CGCCAACACC	TGCGCCAAAG
	1101	CCTGCTTGAA	ACACGGGAAC	ACAGTTGA		

This encodes a protein having amino acid sequence [<SEQ ID 634>] (SEQ ID NO: 634):

	1	MNTSQRNRLV	SRWLNSYERY	RYRRLIHAVR	LGGAVLFATA	SARLLHLQHG
40	51	EWIGMTVFVV	LGMLQFQAI	YSKAVERMLG	TVIGLGAGLG	VLWLNQHYFH
	101	GNLLFYLTVG	TASALAGWAA	VGKNGYVPM	AGLTMCM	DNGSEWFD
	151	LMRAMNVLIG	AAIAIAAAKL	LPLKSTLMWR	FMLADNL	SKMIAEIS
	201	RRMTRERLEE	NMAKMRQINA	RMVKSRS	ATSGESR	AMMEAMQ
	251	RKIVNTTELL	LTTAAKLQSP	KLNGSEIRLL	DRHFTLLQ	LQQTVALIN
	301	RHARRIRIDT	AINPEALEA	EHLHYQWQGF	LWLSTNM	ISALVILLQ
45	351	TRRKWLDAHE	ROHLRQSLLE	TREHS*		

ORF146a (SEQ ID NO: 634) and ORF146-1 (SEQ ID NO: 632) show 99.5% identity in 374 aa overlap:

50 orf146a.pep MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
orf146-1 MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV

	orf146a.pep	LGMLQFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVGTSALAGWAA
	orf146-1	LGMLQFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVGTSALAGWAA
5	orf146a.pep	VGKNGYVPMLAGLTMCM LIGDNGSEWFD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
	orf146-1	VGKNGYVPMLAGLTMCM LIGDNGSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
	orf146a.pep	FMLADNLTDCSKMIAEISNGRRMTREERLEENMAKMRQINARMVKSRSHLAATSGESRISP
	orf146-1	FMLADNLADCSKMIAEISNGRRMTREERLEENMAKMRQINARMVKSRSHLAATSGESRISP
10	orf146a.pep	AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
	orf146-1	AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
	orf146a.pep	RHARRIRIDTAINPELEALAEHLHYQWQGFLWLSTNMQRQISALVILLQRTTRRKWLDAHE
15	orf146-1	RHARRIRIDTAINPELEALAEHLHYQWQGFLWLSTNMQRQISALVILLQRTTRRKWLDAHE
	orf146a.pep	RQHLRQSLLETREHSX
	orf146-1	RQHLRQSLLETREHGX

Homology with a predicted ORF from *N.gonorrhoeae*

- 20 ORF146 (SEQ ID NO: 630) shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) (SEQ ID NO: 636) from *N.gonorrhoeae*:

	orf146.pep	RHARRIRIDTAINPELEALAEHLHYQWQGF	30
	orf146ng	KLNGSEIRLLDRHFTLLQTDLQQTAAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF	364
25	orf146.pep	LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHLRQSLLETREHG	75
	orf146ng	LWLSTNMQRQEISALVIPLQRTTRRKWLDAHERQHLRQSLLETREHG	409

- 30 An ORF146ng nucleotide sequence [<SEQ ID 635>] (SEQ ID NO: 635) was predicted to encode a protein having amino acid sequence [<SEQ ID 636>] (SEQ ID NO: 636):

	1	MSGVRFPSPA	PIPSTDPPSG	SLCFFTFPLQ	TASDMNSSQR	KRLSGRWLNS
	51	YERYRHRLI	HAVRLGGTVL	FATALARLLH	LQHGEWIGMT	VFVVLGMLQF
	101	QGAIYSNAVE	RMLGTVIGLG	AGLGVLWLNQ	HYFHGNLLFY	LTIGTASALA
	151	GWAAVGKNGY	VPMLAGLTM	MLIGDNGSEW	LDSGLMRAMN	VLIGAAIAIA
35	201	AAKLLPLKST	LMWRFMLADN	LADCSKMIAE	ISNGRRMTRE	RLEQNMVKMR
	251	QINARMVKSR	SHLAATSGES	RISPSMMEAM	QHAHRKIVNT	TELLLTAAK
	301	LQSPKLNSE	IRLLDRHFTL	LQTDLQQTAA	LINGRHARRI	RIDTAINPEL
	351	EALAEHLHYQ	WQGFLWLSTN	MRQEISALVI	PLQRTTRRWL	DAHERQHLRQ
40	401	SLLETREHG*				

Further work revealed the following gonococcal DNA sequence [<SEQ ID 637>] (SEQ ID NO: 637):

1 ATGAACTCCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
 51 CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCgga
 101 ccgtCCTGTT CGCCACCGCA CTCGCCGgc tACTCCACCT CCAacacggc
 151 gAATGGATAG GGAtgaCCGT CTTCGTCGTC CTCGGCATGC TCCAGTTCCA
 5 201 AGGCgcgatt tActccaacg cgggtgGAacg taTGctcggg acgggtcatcg
 251 ggctgGGCGC GGGTTTGGgc gTTTATGGC TGAACCAGCA TTAttccac
 301 ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccgg
 351 ctGGGCGGCG GTCGGCAAAA acggctacgt ccctatgctg GCGGGGctgA
 401 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
 10 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC
 501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGTATGA CGCGCGAACG TTTGGAGCAG AATATGGTCA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCACCTCGCC GCCACATCGG
 15 701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAAGT
 801 GCAATCTCCC AAACTCAACG GCAGCGAAAT CCGGCTGCTC GACCGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 20 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

25 This corresponds to the amino acid sequence [SEQ ID 638; ORF146ng-1] (SEQ ID NO: 638; ORF146ng-1):

1 MNSSQRKRLS GRWLSYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG
 51 EWIGMTVFVV LGMLQFQGA IYNAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLDSG
 151 LMRAMNVLIG AAIATAAKL LPLKSTLMWR FMLADNLAD SKMIAEISNG
 201 RRMTRERLEQ NMVKMRQINA RMVKSRS HLAATSGESRISP SMMEAMQHAH
 251 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTAAALING
 301 RHARRIRIDT AINPELEALA EHLHYQWGF LWLSTNM RQEISALVILLQ
 351 TRRKWLDAHE RQHLRQSLLE TREHG*

ORF146ng-1 (SEQ ID NO: 638) and ORF146-1 (SEQ ID NO: 632) show 96.5% identity in 375 aa overlap

orf146-1.pep MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
 40 orf146ng-1 MNSSQRKRLSGRWLSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV
 orf146-1.pep LGMLQFQGA IYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIVGTASALAGWAA
 orf146ng-1 LGMLQFQGA IYNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
 45 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAKLLPLKSTLMWR
 orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAKLLPLKSTLMWR
 orf146-1.pep FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRS HLAATSGESRISP
 orf146ng-1 FMLADNLADCSKMIAEISNGRRMTRERLEQN MVKMRQINARMVKSRS HLAATSGESRISP
 50 orf146-1.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING

```

      :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
orf146ng-1  SMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTAAALING

orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGFLLWLNMRQEISALVILLQTRRKWLDAHE
5 orf146ng-1  RHARRIRIDTAINPELEALAEHLHYQWQGFLLWLNMRQEISALVILLQTRRKWLDAHE

orf146-1.pep RQHLRQSLLTREGHX
orf146ng-1    RQHLRQSLLTREGHX

```

10 Furthermore, ORF146ng-1 (SEQ ID NO: 638) shows homology with a hypothetical *E.coli* protein (SEQ ID NO: 1150):

```

15 sp|P33011|YEEA ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION
   )gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:o348#20; similar to [SwissProt
   Accession Number P33011] [Escherichia coli] )gi|1736682|gnl|PID|d1016560 (D90839)
   ORF_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]
   )gi|1788318 (AE000292) f352; 100% identical to fragment YEEA_ECOLI SW: P33011 but
   has 203 additional C-terminal residues [Escherichia coli] Length = 352
   Score = 109 bits (271), Expect = 2e-23
   Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

20 Query: 20 YRHRRLIHAVRLGGTVLFFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAIYSNAVERML 79
      YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
Sbjct: 15 YRHYRIVHGTTRVALAFLLTFLIIRLFTIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74

Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGKNGYVPMLAGLTMCMCLI 139
      GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++
25 Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

Query: 140 GDNGSEWLD SGLMRAMNVLIGXXXXXXXXXXKLPLKSTLMWRFMLADNLADCSKMIABEISN 199
      G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +
Sbjct: 132 GSPTGE-IDTALWRS G DVILGSL LAMLFTGIWPQRAFIHWRIQLAKSLTEYNRVYQSAFS 190

Query: 200 GRRMTREERLEQN MVKMRQINARMVKSRSRLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259
      + R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V
30 Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRI PKSIYEGIQ TINRNLVCMLEL 247

Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXXAALINGRHARRIRIDTAINPEL 316
      + LN ++R D AL G +N +
Sbjct: 248 QINAYWATRP SHFVLLNAQKLR--DTQHMMQ QILLSLVHALYEGNPQPVFANTEKLND AV 305

35 Query: 317 EALAEHL--HYQWQ-----GFLWLNMRQEISALVILLQTRRK 354
      E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLLNHHDLKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352

```

On the basis of this analysis, including the identification of several transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 639>] (SEQ ID NO: 639)

```
1  .GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA
5  51  GGGCAAAC TC GTCAGTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA
101 AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTTCCGAT
151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAAACTCG CCCGCCGCGT
201 GCGTGAGGCC GGGTTTAAAG TCGTTCCTCGT CGTGGGCGCA AC.GCGGTGA
251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC
301 GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAAACTGT TTGCCAAATG
10 351 GGTGCGGGCG GCGTTTCCTA TCGTCATGTT TGAAACGCCG CACCGCATCG
401 GTGCAGCGCT TGCCGATATG GCGGAACGTG TCCCCGAACG CCGATTAATG
451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCCTAA GCGGCACGCT
501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG
551 AGATGGTGTT GGTGCTTTAT CCGGCGCAGG ATGAAAAACA CGAAGGCTTG
15 601 TCCGAGTCCG CGCAAAACAT CATGAAATC CTCACAGCCG AGCTGCCGAC
651 CAAACAGGCG GCGGAGCTTG CTGCCAAAT CACGGGCGAG GGAAAGAAAG
701 CTTTGTACGA T..
```

This corresponds to the amino acid sequence [<SEQ ID 640; ORF147>] (SEQ ID NO: 640; ORF147):

```
1  .AEDTRVTAQL LSAYGIQGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51  AGTPAVCDPG AKLARRVREA GFKVVPVVG XAVMAALSVA GVEGSDFYFN
101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGAALADM AELFPERRLM
151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL
25 201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKKALYD..
```

Further work revealed the complete nucleotide sequence [<SEQ ID 641>] (SEQ ID NO: 641):

```
1  ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
30 51  ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CCGCTTACCG CACAGCTTTT GAGCGCGTAC GGCATTGAGG GCAAACCTCGT
201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGGT
251 ATCTTTTACA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGGCG GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
35 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGGC
501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
551 CCATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
40 601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAACATCA TGAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801 GGAGCTTGCT GCCAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
45 851 TGGCTCTGTC TTGAAAAAC AAATAG
```

This corresponds to the amino acid sequence [<SEQ ID 642; ORF147-1>] (SEQ ID NO: 642; ORF147-1):

```
1  MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
```

51 RVT AQLLSAY GIQGLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLMLARE
 201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 (SEQ ID NO: 1151) of *E.coli* (accession number U18997)

10 ORF147 (SEQ ID NO: 640) and *E.coli* ORF286 protein (SEQ ID NO: 1151) show 36% aa identity in 237aa overlap:

Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
 AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
 Orf286: 43 AEDTRHTGLLLQHFGINARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPG 102
 15 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLF AKWVRA 120
 L R RE F + GF+P KS RR
 Orf286: 103 YHLVRTCREAGIRVVPLPGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAI EAE 162
 Orf147: 121 AFPIVMFETPHRIG AALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179
 ++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
 20 Orf286: 163 PRTLIFYESTHRL LDSLEDIVAVLGESRYVVLARELTKTWETIHGAPV GELLAWVKEDEN 222
 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELAAKITGEGKKALY 236
 + +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
 Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAELPLKKAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF147 (SEQ ID NO: 640) shows 96.6% identity over a 237aa overlap with ORF75a (SEQ ID NO: 290) from strain A of *N. meningitidis*:

30 orf147.pep AEDTRVTAQLLSAYGIQGLVSVREHNERQ
 orf75a TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGLVSVREHNERQ
 20 30 40 50 60 70
 35 orf147.pep MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA
 orf75a MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVGFVVPVVGASAVMAALSVA
 80 90 100 110 120 130
 40 orf147.pep GVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFPIVMFETPHRIG AALADMAELFPERRLM
 orf75a GVAGSDFYFNGFVPPKSGERRKLF AKWVRVAFPVVMFETPHRIGATLADMAELFPERRLM

-462-

		140	150	160	170	180	190	
			160	170	180	190	200	210
5	orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLVPAQDEKHEGLSESAQNIMKI						
	orf75a	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLVPAQDEKHEGLSESAQNIMKI						
		200	210	220	230	240	250	
			220	230				
	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD						
10	orf75a	LTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX						
		260	270	280	290			

ORF147a is identical to ORF75a (SEQ ID NO: 290), which includes aa 56-292 of ORF75 (SEQ ID NO: 286).

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF147 (SEQ ID NO: 640) shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) (SEQ ID NO: 644) from *N. gonorrhoeae*:

	orf147.pep	AEDTRVTAQLLSAYGIQGKLVSVREHNERQ	30
20	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGRVSVREHNERQ	85
	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA	90
	orf147ng	MADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA	145
25	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPVVFETPHRIGALADMAELFPERRLM	150
	orf147ng	GVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVFETPHRIGATLADMAELFPERRLM	205
	orf147.pep	LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLVPAQDEKHEGLSESAQNIMKI	210
	orf147ng	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLVPAQDEKHEGLSESAQNAMKI	265
30	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD	237
	orf147ng	LAAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF147ng nucleotide sequence [<SEQ ID 643>] (SEQ ID NO: 643) was predicted to encode a protein having amino acid sequence [<SEQ ID 644>] (SEQ ID NO: 644):

35	1	MSVFQTAFFM	FQKHLQKASD	SVVGGTLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTR	VTAQLLSAYG	IQGRVSVRE	HNERQMADKV	IGFLSDGLVV
	101	AQVSDAGTPA	VCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSVAGVAES
	151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP
	201	ERRMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLVPAQDE
40	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
	301	*				

Further work revealed the following gonococcal DNA sequence [SEQ ID 645] (SEQ ID NO: 645):

```

      1  ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
      51  ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCAGAC ATTACCCTGC
5      101  GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATTGTGTC CGAAGACACG
      151  CGCGTTACTG CGCAGCTTTT GAGCGCGTAC GGCATTTCAGG GCAGGTTGGT
      201  CAGATGTGCG GAACACAACG AGCGGCAGAT GGCGGACAAG GTAATCGGTT
      251  TCCTTTTCAGA CGGCCTGGTT GTGGCGCAGG TTCCGATGC GGGTACGCCG
      301  GCCGTGTGCG ACCCGGGGCG GAAACTCGCC CGCCGCGTGC GCGAAGCAGG
      10  351  GTTCAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTAATG GCGGCGTTGA
      401  GTGTGGCCGG TGTGGCGGAA TCCGATTTTT ATTCAACGG TTTGTACCG
      451  CCGAAATCGG GCGAACGTAG GAAATGTTT GCCAAATGGG TCGCGGCGGC
      501  ATTTCTGTGC GTCATGTTTG AAACGCCGCA CCGAATCGGG GCAACGCTTG
      551  CCGATATGGC GGAATTGTTT CCCGAACGCC GTCTGATGCT GGCGCGCGAA
      15  601  ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
      651  GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
      701  TGCTTTATCC GCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCTGCG
      751  CAAAATGCGA TGAATCCTT TCGCGCCGAG CTGCCGACCA AGCAGGCGGC
      801  GGAGCTTGCC GCCAAGATTA CAGGTGAGGG CAAAAGGCT TTGTACGATT
      20  851  TGGCACTGTC GTGAAAAAC AAATGA

```

This corresponds to the amino acid sequence [SEQ ID 646; ORF147ng-1] (SEQ ID NO: 646; ORF147ng-1):

```

      1  MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRLALAVLQ KADIICAEDT
      25  51  RVTAQLLSAY GIQRLVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
      101  AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
      151  PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRMLLARE
      201  ITKTFFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
      251  QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

ORF147ng-1 (SEQ ID NO: 646) shows homology to a hypothetical *E.coli* protein (SEQ ID NO: 1152):

```

35  sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    )gi|606086 (U18997) ORF_f286 [Escherichia coli]
    )gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
    Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

40  Query: 4  KHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
      K  Q  A  +S  G  LY+V TPIGNLADIT RAL VLQ  D+I AEDTR T  LL  +GI
      Sbjct: 2  KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAEDTRHTGLLLQHFGIN 59

      Query: 64  GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
      RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
45  Sbjct: 60  ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

      Query: 124  VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKWVRAAFPVVMFETPHRIGATL 183
      G  A  + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
      Sbjct: 120  PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAI EAEPRTLIFYESTHRLLDL 179

```

Query: 184 ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
 D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
 Sbjct: 180 EDIVAVLGESEYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

Query: 243 HEGLSESAQNAMKILAAELPTKQAAELAALKITGEGKKALYDLAL 286
 E L A + +L AELP K+AA LAA+I G K ALY AL
 Sbjct: 239 EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 647>] (SEQ ID NO: 647)

```

15      1  ATGAAAACAA CCGACAAACG GACAAACGAA ACACACCGCA AAGCCCCGAA
      51  AACCGGTCGC ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
     101  TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
     151  TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
     201  GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
     251  CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
     301  GTGGCGGcAT TGGTGGGCGt ATCAATATAT TGTGAGCGTG GCACATAACG
     351  GCGGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAak.AA tATCCC.GAT
     401  CAACAwCGww TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
     451  GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGCATA
     501  AATwTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG
     551  CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTT GTATTGGGGC
     601  AGGCAGGCAA TATTGCGGAT CTGATGAAGA TGAGCCCAAT AACC CGGAAA
     651  GTTCATATCA TATTGCAAGT .....
     701  ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
     751  AAAGTGTTA ATTAATGGGG TATTGCAAAAC GGGCAACCCC TATATAGGAA
     801  AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
     851  TTTGCTGAG ATACCCATTG AGTATTCTAC GAACCACTG CAAAATGGGAA
     901  ATACTCTTTT AACGACGATA ATAATGCGAC AGGAAAAATC AATGCCAAAC
     951  ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG
    1001  TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC
    1051  AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAAATATTT
    1101  CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT
    1151  CAAGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA
    1201  AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA
    1251  CCGTTACTTG GAAAGTAAAC GGCGTGCGCA ACGACCGCCT GTCCAAAATC
    1301  GGCAAAGGCA CGCTG.....

//

2101  ..... GATAAAG
2151  TGA CTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC
2201  GATCAGCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT
2251  TAGTGCAAAT GCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA
2301  ACGGCAACCK TAGCCTCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC
2351  ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG
2401  CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCCGGC AACGCTAAGG
2451  CAAACGTAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG
2501  GCAGTATTCC ATTTTGAAAAG CAGCCGCTTT ACCGGACAAA TCAGCGGCGG

```


5
10
15
20
25

```

2551 CAagGATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCCGTCag
2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT
2651 TCCGCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGa
2701 TGCGCCGCGC CGCCGTTTCGC GCCGTTTCGC CCGTTCCCTA TTATmCGTTA
2751 CACCGCCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAACGGC
2801 AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA
2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT
2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG
2951 GTAGTGGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC
3001 CCTGCAAAAC GAACACGTCG ATGCAGGCGC GTGG.....

//

3551 ..... TTAGAC CGCGTATTTG CCGAAGACCG
3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAAA CACTACCGTT
3651 CGCAAGATTT CCGCGCTAC CGCCAACAAA CCGACCTGCG CCAAAATCGGT
3701 ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGCATCCTGT TTTCGCACAA
3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG
3801 CCCACGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC
3851 ATCAGnCGCG GCGCGGGT TTAGCAGCGG CAGCCTTTca GACGGCATCG
3901 GAGsmAAAwT CCGCCGCGC GTGctGCATT ACGGCATTCA GGCACGAtAC
3951 CGCGCCGgtt tCgGcGgAtT CCGCATCGAA CCGCACATCG GCGCAACGcG
4001 ctATTTTCGTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA
4051 CCCCCGGCCT TGCATTCAAC CgcTACCGCG CGGGCATTaa GGCAGATTAT
4101 TCATTCAAAC CGGCGCAACA CATTTCATC ACGCCTTATT TGAGCCTGTC
4151 CTATACCGAT GCCGCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCCG
4201 TATTGGCTCA GGATTTCGGC AAAACCCGCA GTGCGGAATG GGgCGTAAAC
4251 GCCGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCAAAGG
4301 CCCGCAACTG GAAGCGCAAC ACAGCGCGG CATCAAATTA GGCTACCGCT
4351 GGTAA...

```

30 This corresponds to the amino acid sequence [<SEQ ID 648; ORF1>] (SEQ ID NO: 648; ORF1):

35
40
45
50
55

```

1 MKTTDKRTE THRKAPKTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDFaEN KGKFAVGAkd IEVYNKKGEL VGKSMTKAPM IDFSVVSrNG
101 VAALVGvQYI VsvAHNGGYN NVDFGAEGXN IXDQXRrTYK IVKRNNYKAG
151 TKGHpYGGDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYpDRVRIGA
201 GRQYrSDED EPNnRESSYH IAS.....GS PMFIYDAQKQ
251 KWLINGVLQT GNPYIGKSNG FQLVRKdWfY DEIFAGDTHS VFYEPRQNGK
301 YSFNDdNNGT GKINAKHEHN SLPnRLKTRT VQLFNvSLSE TAREPVYHAA
351 GGVNSYRpRL NNGENISFID EGKGELILTS NINQAGGLY FQGDFTVSPE
401 NNETWQgAGV HISEDSTvTW KVNGVANDRL SKIGKGTl...

//

701 .....DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL
751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS
801 DHAVQNGSLT LSGNAKANVS HSALNGNVSL ADKAVFHFES SRFTGQISGG
851 KDTALHLKDS EWTLPsGXEL GNLNLdNATI TLNSAYRHDA AGAQTSATD
901 APRRRSRRSR RSLlXvTPPT SVESRFNTLT VNGKLNGQGT FRFMSElFGY
951 RSDKLKLAEs SEGTYTLAVN NTGNEPASLE QLTvVEGKDN KPLSENlNFT
1001 LQNEHVDAGA W.....

//

1151 ..... LDRVFAEDR
1201 RNAVWTSGIR DTKHYRSQDF RAYRQQTDLR QIGMQKNLGS GRVGILFSHN
1251 RTENTFDdGI GNSARLAHGA VFGQYgIDRF YIGISAGAGF SSGSLSDGIG
1301 XKXRRRVLHY GIQARYRAGF GGFGIEPHIG ATRYFvQKAD YRYENvNIAT
1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDAASG KVRTRvNTAV
1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW
1451 *

```

Further sequencing analysis revealed the complete nucleotide sequence [<SEQ ID 649>] (SEQ ID NO: 649):

5
10
15
20
25
30
35
40
45
50
55
60

1 ATGAAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCGT
101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
401 ATCGTTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT
451 AAAGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCATAAAAT
501 TGTACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCGTAT TGGGGCAGGC
601 AGGCAATATT GGCATCTGA TGAAGATGAG CCAATAAACC GCGAAAGTTC
651 ATATCATATT GCAAGTGCCT ATTCTTGGCT CGTTGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGAAAAAATT
751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGCAG
801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
851 ATGGGGTATT GCAAACGGGC AACCCTATA TAGGAAAAAG CAATGGCTTC
901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTTCAGTA TTCTACGAAC CACGTCAAAA TGGGAAATAC TCTTTTAACG
1001 ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT
1051 CTGCCAATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTCT TATTGACGAA
1201 GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
1251 ATTATATTTT CAAGGAGATT TTACGGTCTC GCCTGAAAAT AACGAACTT
1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGAAG ACAGTACCGT TACTTGGAAA
1351 GTAAACGGCG TGGCAAACGA CCGCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
1451 GTACAGTCAT TTTGATCAG CAGGCAGACG ATAAAGGCAA AAAACAAGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGTACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCGGCTTT CGCGGCGGAC
1601 GTTTGGATTT AAACGGGCAT TCGTTTCGT TCCACCGTAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACAAACAGT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAGC CCGCCGAGA
1851 AGACCGCACC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA
1901 CGCAAAACAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
1951 TACAATCATT TAAACGACCA TTGGTCGCAA AAAGAGGGCA TTCTCGCGG
2001 GGAAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAAGCGGAAA
2051 ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCGCAA TGTGCCCCAA
2101 GTGAAAGCG ATTTGGCATT GAGCAATCAC GCCCAAGCAG TTTTGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAATTG TGTCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTGCCG ATCAGCTCA
2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAAATG
2351 GCGATACACG TTATACAGTC AGCCACAACG CCACCCAAAA CGGCAACCTT
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAACCG
2451 CAACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
2501 TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC
2551 CATTCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTACAG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCTATCG
2751 CCACGATGCG GCAGGGGCGC AAACCGGCAG TCGACAGAT GCGCCGCGC
2801 GCCGTTTCGCG CCGTTCGCGC CGTTCCTTAT TATCCGTTAC ACCGCCAACT
2851 TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
2901 TCAGGGAACA TTCCGCTTTA TGTGGAACCT CTTGCGCTAC CGCAGCGACA
2951 AATTGAAGCT GCGGAAAGT TCCGAAGGCA CTTACACCTT GGCGGTCAAC
3001 AATACGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGAAGG

5
10
15
20
25

```

3051 AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTACCC CTGCAAAACG
3101 AACACGTCGA TGCCGCGCGG TGGCGTTACC AACTCATCCG CAAAGACGGC
3151 GAGTTCCGCC TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAAACT
3201 CGGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAC GCGCAAAGCC
3251 TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC
3301 GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTCT GCATTATGCA
3351 GGGCGAGGAA GAGAAAAAAC GGGTGAGGCG GGATAAAGAC ACCGCCTTGG
3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTTCCCCCGC
3451 GCCCGCCGCG CCCCGCGGGA TTTGCGCAA CTGCAACCCC AACCGCAGCC
3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTTGAGTG
3551 AATTTTCCGC CACGCTCAAC AGCGTTTTCG CCGTACAGGA CGAATTAGAC
3601 CGCGTATTTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651 GGACACCAAA CACTACCGTT CGCAAGATTT CCGCGCCTAC CGCCAACAAA
3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC
3751 GGCATCCTGT TTTTCGACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
3801 CCGCAACTCG GCACGCTTG CCCACGCGCG CGTTTTCGGG CAATACGCA
3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC
3901 AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCCGCCGCG TGCTGCATTA
3951 CGGCATTGAG GCACGATACC GCGCCGTTT CCGCGGATTG GGCATCGAAC
4001 CGCACATCGG CGCAACGCGC TATTTCTGTC AAAAAGCGGA TTACCGCTAC
4051 GAAAACGTCA ATATCGCCAC CCCCAGCCTT GCATTCAACC GCTACCGCGC
4101 GGGCATTAAAG GCAGATTATT CATTCAAACC GGCGCAACAC ATTTCCATCA
4151 CGCCTTATTT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201 ACACGCGTCA ATACCGCGCT ATTGGCTCAG GATTTTCGGA AAACCCGCAG
4251 TGCGGAATGG GCGGTAAACG CCGAAATCAA AGGTTTCACG CTGTCCCTCC
4301 ACGCTGCCGC CGCCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351 ATCAAATTAG GCTACCGCTG GTAA

```

This corresponds to the amino acid sequence [<SEQ ID 650; ORF1-1>] (SEQ ID NO: 650; ORF1-1):

30
35
40
45
50
55

```

1  MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51  YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG
101 VAALVGDQYI VSVAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNRYKAGT
151 KGHYPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
201 RQYWRSDEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
251 KHSPYGFLLPT GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNGF
301 QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS
351 LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE
401 KGKELILTSN INQAGAGLYF QGDFTVSPEN NETWQAGAVH ISEDSTVTWK
451 VNGVANDRLS KIGKGTLLHVQ AKGENQGSIS VGDGTVILDQ QADDKGGKQA
501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
551 DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGKED
601 TTKTNGRLNL VYQPAAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
701 VKGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
801 SLVNGAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
851 HSLAGNVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSTEL
901 GNLNLNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
951 SVESRFNTLT VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN
1001 NTGNEPASLE QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLRKDG
1051 EFRLHNPVKE QELSDKLKGA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101 VAEPARQAGG ENVGIMQAE EKKRVQADKD TALAKQREAE TRPATTAFFR
1151 ARRARRDLPO LQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQSDLD
1201 RVFAEDRRNA WPTSGIRDTK HYRSQDFRAY RQQTDLRQIG MQKNLQSGRV
1251 GILFSHNRTN NTFDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
1301 SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAAAGKVR

```

1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
1451 IKLGYRW*

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 5 ORF1 (SEQ ID NO: 648) shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) (SEQ ID NO: 652) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf1.pep	MKT	TDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGIN	YQYYR	DAEN		
10	orf1a	MKT	TDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGIN	YQYYR	DAEN		
		10	20	30	40	50	60
	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSM	TAPMIDFSVSRNGVAALVG	VQYIVSVAHNGGYN			
15	orf1a	KGKFAVGAKDIEVYNKKGELVGKSM	TAPMIDFSVSRNGVAALVG	DQYIVSVAHNGGYN			
		70	80	90	100	110	120
	orf1.pep	NVDFGAEGXNIXDQXRTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKKXVTDAEPVEMTSY					
20	orf1a	NVDFGAEGXN-PDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTDAEPVEMTSD					
		130	140	150	160	170	180
	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSD	ED	EP	-----	NN	----
25	orf1a	MRGNTYS	DK	KEYP	PERVRIGSGHHYWR	YDDDKHGDLSYSGAWLIGGNTHMQGWNNGVXSL	
		180	190	200	210	220	230
	orf1.pep	----	RESSYH	----	IA	----	SGSPMFIYDAQKQKWLINGVLQTGNPNYIGKSNGFQLVRK
30	orf1a	SGDVRHANDYGPMPIAGAAGDSGSPMFIYDKTNKWLINGVLQTGYPSGRENGFQLIRK					
		240	250	260	270	280	290
	orf1.pep	DWFYDEIFAGDTHSVFYEP	RQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTVQLFNV				
35	orf1a	DWFYDDIYRGDTHTVXFEP	RSNGHFSFTSNNGTGTVTETNEKVSNP-KLKVQTVRLFDE				
		300	310	320	330	340	350
	orf1.pep	SLSETAREPVYHAAGGVNSYR	PRLNNGENISFIDEKG	GELILTSNINQGAGGLYFQGDFT			
40	orf1a	SLNETDK	KEPVY-AAGGVNQYRPRLNNGENLSFIDYGN	GKLILSNINQGAGGLYFEGDFT			
		360	370	380	390	400	410

5	orfl1.pep	VSPENNETWQAGVHI	SEDSTVTWKVNGV	ANDRLSKIGKGT	-----	
	orfla	VSPENNETWQAGVHI	SEDSTVTWKVNGV	ANDRLSKIGKGT	LHVQAKGENQGSISV	GDGT
10	orfl1.pep	-----				
	orfla	VILDQQADDKGGKQAF	SEIGLXSGRGTVQL	NADNQFNPDKLYFG	FRGGRLDLNGHSLSF	H
15	orfl1.pep	-----				
	orfla	RIQNTDEGAMIXXHN	ATTTSTVTITGNES	ITQPSGKNINRLNYS	KEIAYNGWFGEKDTT	K
20	orfl1.pep	-----				
	orfla	TNGRLNLVYQPAAED	RTXLLSGGTNLNGN	ITQTNGKLFFSGRPT	PHAYNHLGSGWSKME	G
25	orfl1.pep	-----				
	orfla	IPQGEIVWDNDWIXR	TFKAENFHIQGGQAV	ISRNVAKVEGDHLSN	HAQAVFGVAPHQSH	
30	orfl1.pep	-----	440	450	460	470 480
	orfla	TICTRSDWTGLTNCV	EXXITDDKVIASLT	KTDXSGXVXLXXXXX	XLGXAXLXGNLSAN	
35	orfl1.pep	490	500	510	520	530 540
	orfla	GDTRYTVSHNATQNG	NXSLVXNAQATFNQ	ATLNGNTSASGNAS	FNLSDAHVQNGSLT	LSG
40	orfl1.pep	550	560	570	580	590 600
	orfla	NAKANVSHSALNGNV	SLADKAVFHFESSR	FTGQISGGKDTALH	LKDSEWTLPSGXEL	GNL
45	orfl1.pep	610	620	630	640	650 660
	orfla	NLDNATITLNSAYRH	DAAGAQTGSATDAP	RRRRSRSSRLXVTP	PPTSVESRFNTLT	VNG
50	orfl1.pep	670	680	690	700	710 720
	orfla	KLNGQGTFRFMSELF	GYRSDKCLKLAES	SEGTYTLAVNNTGN	EPASLEQLTVVEG	KDNKPL

-470-

		730	740	750	
	orf1.pep	SENLNFTLQNEHVDAGAW-----			
	orf1a	SENLNFTLQNEHVDAGAWRYQLIRKDGFEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQS			
5		1020	1030	1040	1050 1060 1070
	orf1.pep	-----			
	orf1a	LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKRVQADKDSALAKQREAETRP			
10		1080	1090	1100	1110 1120 1130
	orf1.pep	-----			
	orf1a	XTTAFPRARXARRDLPPQPQPQPQPQDLXSRYANSGLSEFSATLNSVFAVQDELDR			
15		1140	1150	1160	1170 1180 1190
	orf1.pep	-----			
	orf1a	VFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGLFSHNRTE			
20		1200	1210	1220	1230 1240 1250
	orf1.pep	-----			
	orf1a	VFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGLFSHNRTE			
	orf1.pep	-----			
	orf1a	TFDDGIGNSARLAHGA VFGQY GIDRFYIGISAGAGFSSGSLSDGIGKXRRRVVLHYGIQA			
25		1260	1270	1280	1290 1300 1310
	orf1.pep	-----			
	orf1a	XFDDGIGNSARLAHGA VFGQY GIGRFDIGISTGAGFSSGXLSDGIGGKIRRRVLHYGIQA			
	orf1.pep	-----			
	orf1a	RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI			
30		1320	1330	1340	1350 1360 1370
	orf1.pep	-----			
	orf1a	RYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX			
	orf1.pep	-----			
	orf1a	SITPYLSLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAKGP			
35		1380	1390	1400	1410 1420 1430
	orf1.pep	-----			
	orf1a	SITPYXLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHAAAAKGP			
	orf1.pep	-----			
	orf1a	QLEAQHSAGIKLGYRWX			
40		1440	1450		

The complete length ORF1a nucleotide sequence [<SEQ ID 651>] (SEQ ID NO: 651) is:

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCGT
45	101	TCGGCATTCT	TCCCCAAGCT	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTNT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
50	351	CGGCTATAAC	AACGTTGATT	TTGGTGCGGA	AGGAAGNAAT	CCCGATCAGC
	401	ACCGTTTTTC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	GCCTGACAAAT

5 451 TCACACCCCTT ACAACGGCGA TTANCATATG CCGCGTTTGC ATAAATTTGT
501 CACAGATGCA GAACCTGTCTG AAATGACGAG TGACATGAGG GGAATACCT
551 ATTCCGATAA AGAAAAATAT CCCGAGCGTG TCCGCATCGG CTCAGGACAC
601 CACTATTGGC GTTATGATGA TGACAAACAC GGCGATTTAT CCTACTCCGG
651 CGCATGGTTA ATTGGCGGCA ATACACATAT GCAGGGTTGG GGAAATAATG
701 GCGTANTTAG TTTGAGCGGC GATGTGCGCC ATGCCAACGA CTATGGCCCT
751 ATGCCGATTG CAGGTGCGGC AGGCGACAGC GGTCGCCAA TGTTTATTTA
801 TGACAAAACA AACAATAAAT GGCTGCTCAA CGGAGTTTTA CAAACCGGCT
10 851 ACCCTTATTC CGGCAGGGAA AACGGTTTCC AGCTGATACG CAAAGATTGG
901 TTCTACGATG ACATTACAG AGGCGATACA CATACCGTCT NTTTTGAACC
951 GCGCAGTAAC GGACATTTTT CCTTTACATC CAACAACAAC GGTACGGGTA
1001 CGGTAACAGA AACCACGAA AAGGTNTCCA ATCCAAAGCT TAAAGTACAG
1051 ACAGTCCGAC TGTTTGACGA ATCTTTGAAT GAACTGATA AAGAACCAGT
1101 TTACGCGGCA GGGGGTGTTA ATCAGTACCG TCCAAGGTTA AACACCGGTG
15 1151 AAAACCTTTC TTTTATCGAT TACGGCAACG GCAAACTCAT CTTATCAAAC
1201 AACATCAACC AAGCGCGGG CGGTTTGAT TTTGAAGGTG ATTTTACGGT
1251 CTCGCCGTGA AACAACGAAA CGTGCAAGG CGCGGGCGTT CATATCAGTG
1301 AAGACAGTAC CGTTACTTGG AAAGTAAACG GCGTGGCAA CGACCGCCTG
1351 TCCAAAATCG GCAAAGGCAC GCTGCACGTT CAAGCCAAAG GGGAAAACCA
20 1401 AGGCTCGATC AGCGTGGGCG ACGGTACAGT CATTTTGGAT CAGCAGGCAG
1451 ACGATAAAGG CAAAAACAA GCCTTTAGTG AAATCGGCTT GNTCAGCGGC
1501 AGGGGTACGG TGCAACTGAA TGCCGATAAT CAGTTCACC CCGACAACT
1551 CTATTTTCGGC TTTTCGCGCG GACGTTTGA TTTAAACGGG CATTCGCTTT
25 1601 CGTTCCACCG TATTCAAAAT ACCGATGAAG GGGCGATGAT TGNCNATCAT
1651 AATGCCACAA CAACATCCAC CGTTACCATT ACAGGGAATG AAAGTATTAC
1701 ACAACCGAGT GGTAAGAATA TCAATAGACT TAATTACAGC AAAGAAATTG
1751 CCTACAACGG TTGGTTTGGC GAGAAAGATA CGACCAAAAC GAACGGGCGG
1801 CTCAACCTTG TTTACCAGCC CGCCGAGAA GACCGCACCC NGCTGCTTTC
1851 CGGCGGAACA AATTAAACG GCAACATCAC GCAACAAAC GGCAAACTGT
30 1901 TTTTCAGCGG CAGACGACA CCGCACGCCT ACAATCATTT AGGAAGCGGG
1951 TGGTCAAAAA TGGAAGGTAT CCCACAAGGA GAAATCGTGT GGGACAACGA
2001 CTGGATCNAC CGCACGTTTA AAGCGGAAAA TTTCCATATT CAGGCGGGG
2051 AGGCGGTGAT TTCCCACAAT GTTGCCAAAG TGGAAGGCGA TTGNCATTTG
2101 AGCAATCACG CCCAAGCAGT TTTTGGTGTC GCACCGCATC AAAGCCATAC
35 2151 AATCTGTACA CGTTCGGACT GGACNGGTCT GACAAATTGT GTCGAANAAA
2201 NCATTACCGA CGATAAAGTG ATTGCTTCAT TGAATAAGAC NGACNTNAGC
2251 GGCANTGTNA GNCTNNCCNA TNACGNTNNT TNAAANCTCN CNGGGCNTGC
2301 NNCACNTAAN GGCAATCTTA GTGCAAATGG CGATACACGT TATACAGTCA
2351 GCCACAACGC CACCCAAAAC GGCAACCTTA GCCTCGTGGG CAATGCCCAA
40 2401 GCAACATTTA ATCAAGCCAC ATTAACGGC AACNCATCGG NTTCCGGCAA
2451 TGCTTCATTT AATCTAAGCA ACAACGCCGC ACAAACGGC AGTCTGACGC
2501 TTTCCGACAA CGCTAAGGCA AACGTAAGCC ATTCCGCACT CAACGGCAAT
2551 GTCTCCCTAG CCGATAAGGC AGTATTCCAT TTTGAAAAA GCGCGTTTAC
2601 CGGACAACTC AGCGGCAGCA AGGANACAGC ATTACACTTA AAAGACAGCG
45 2651 AATGGACGCT GCCGTACGGC ACGGAATTAG GCAATTTAAA CCTTGACAAC
2701 GCCACCATTA CACTCAATTC CGCCTATCGC CACGATGCTG CAGGCGCGCA
2751 AACC GGCAAGT GTGTACAGCA CGCCGCGCCG CCGTTCGCGC CGTTCCTTAT
2801 TATCCGTTAC ACCGCCAAT TCGGTAGAAT CCCGTTTCAA CACGCTGACG
2851 GTAAACGGCA AATTGAAACNG TCAAGGAACA TTCCGCTTTA TGTCCGAACT
50 2901 CTTCCGCTAC CGAAGCGACA AATTGAAGCT GGCGGAAAGT TCCGAAGGNA
2951 CTTACACCTT GGCGGTCAAC AATACCGGCA ACGAACCCGT AAGCCTCGAT
3001 CAATTGACGG TAGTGGAAGG GAAAGACAAC AAACCGCTGT CCGAAAACCT
3051 TAATTTACAC CTGCAAAACG AACACGTCGA TGCCGCGCGG TGGCGTTACC
3101 AACTCATCCG CAAAGACGGC GAGTTCGCGC TGCATAATCC GGTCAAAGAA
55 3151 CAAGAGCTTT CCGACAACT CCGCAAGGCA GAAGCCAAA AACAGGCGGA
3201 AAAAGACAAC GCGCAAGGCC TTGACGCGCT GATTGCGGCC GGGCGCGATG
3251 CCGCCGAAAA GACAGAAAGC GTTGCCGAAC CGGCCCGGCN GGCAGGCGGG
3301 GAAAATGTCG GCATTATGCA GGCGGAGGAA GAGAAAAAAC GGGTGCAGGC
3351 GGATAAAGAC AGCGCNTTGG CGAAACAGCG CGAAGCGGAA ACCCGGCGCG
60 3401 NTACCACCGC CTTCCCCCGC GCGCGNCGC CCCGCCGGA TTTGCCGCAA
3451 CCGCAGCCCC AACC GCAACC TCAACCCCAA CCGCAGCGCG ACCTGATNAG

5
10
15

```

3501 CCGTTATGCC AATAGCGGTT TGAGTGAATT TTCCGCCACG CTCAACACGG
3551 TTTTCGCCGT ACAGGACGAA TTGGACCGCG TGTTTGCCGA AGACCGCCGC
3601 AACGCNGTTT GGACAAGCNG CATCCGNGAC ACCAAACACT ACCGTTTCGA
3651 AGATTTCGCG GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC
3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTT GCACAACCGG
3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTTCGAC ATCGGCATCA
3851 GCACGGGCGC GGGTTTTCAG AGCGGCANTC TNTCAGACGG CATCGGAGGC
3901 AAAATCCGCC GCCCGGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC
3951 CGGTTTCGGC GGATTCGGCA TCGAACCGTA CATCGGCGCA ACGCGCTATT
4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTTCATT
4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCCTATA
4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG
4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGGC GAATGGGGCG TAAACGCCGA
4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGCC AAAGNCCGC
4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA

```

This encodes a protein having amino acid sequence [SEQ ID 652] (SEQ ID NO: 652):

20
25
30
35
40
45

```

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS RNG
101 VAALVGQYI VSVAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN
151 SHPYNGDXHM PRLHKFVTDA EPVEMTSDMR GNTYSDKEY PERVRIGSGH
201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP
251 MPIAGAGDS GSPMFIYDKT NNKWLNLGV L QTGYPYSGRE NGFQLIRKDW
301 FYDDIYRGDT HTVXFEPERSN GHFSFTSNNN GTGTVTETNE KVSNNPKLVQ
351 TVRLFDESLN ETDKEPVYAA GGVNQYRPR L NNGENLSFID YGNGLILSN
401 NINQGAGGLY FEGDFTVSPE NNETWQAGAV HISEDSTVTW KVNGBANDRL
451 SKIGKGLHV QAKGENQGS I SVGDGTVILD QQADDKGGKQ AFSEIGLXSG
501 RGTVQLNADN QFNPDKLYFG FRGGRDLNG HSLSFHRIQ TDEGAMIXXH
551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
601 LNLVYQPAE DRTXLLSGGT NLNGNITQTN GKLFPSGRPT PHAYNHLGSG
651 WSKMEGIPQG EIVWDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGDHXL
701 SNHAQAVFGV APHQSHICT RSDWTGLTNC VEXXITDDKV IASLTKTDXS
751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATQN GNLSLVGNAQ
801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNGN
851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDEWTLPSG TELGNLNLND
901 ATITLNSAYR HDAAGAQTGX VSDTPRRRSR RSLSVTPPT SVESRFNTLT
951 VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
1001 QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG EFRLHNPVKE
1051 QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAAEKTES VAEPARXAGG
1101 ENVGIMQAE EKKRVQADKD SALAQREAE TRPXTTAFPR ARXARRDLPO
1151 PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
1201 NAVWTSXIRX TKHYRSQDFR AYRQQTDLRQ IGMQKNLGS RVGILFVSHNR
1251 TENXFDDGIG NSARLAHGAV FGQYIGRFD IGISTGAGFS SGXLSDGIGG
1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENVNIATP
1351 GLAFNRIRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL
1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAAAA KGPQLEAQHS AGIKLGYRW*

```

50 A transmembrane region is underlined.

ORF1-1 (SEQ ID NO: 650) shows 86.3% identity over a 1462aa overlap with ORF1a (SEQ ID NO: 652):

		10	20	30	40	50	60
	orf1a.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYYQYRDAEN					
5	orf1-1	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYYQYRDAEN					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf1a.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN					
10	orf1-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN					
		70	80	90	100	110	120
		130	140	150	160	170	179
	orf1a.pep	NVDFGAEGXNPQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTDAEPVEMTSDM					
15	orf1-1	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSYM					
		130	140	150	160	170	180
		180	190	200	210	220	230
	orf1a.pep	RGNTYSDEKEYPERVIRIGSGHHYWRYYDDDKHGDL--SYSGA---WLIGGNTMHQGWGNN					
		: : : : : : : : : : : : : : : : : : :					
20	orf1-1	DGRKYIDQNNYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG					
		190	200	210	220	230	240
		240	250	260	270	280	290
	orf1a.pep	GVXSLSGD-VRHANDYGPMPIAGAAGDSGSPMFIYDKTNKWLNLGVLQTGPYPYSGRENG					
		: : : : : : : : : : : : : : : : : : : :					
25	orf1-1	GTVNLGSEKIKHS-PYGFLPTGGSGFGDSGSPMFIYDAQKQKWLINGVLQTGNPNYIGKSNG					
		250	260	270	280	290	
		300	310	320	330	340	350
	orf1a.pep	FQLIRKDWFYDDIYRGDTHTVXFEPRSNHGHSFTSNNGTGTVTETNEKVSNP-KLKVQT					
30	orf1-1	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT					
		300	310	320	330	340	350
		360	370	380	390	400	410
	orf1a.pep	VRLFDESINETDKEPVY-AAGGVNQYRPRLNNGENLSFIDYGNGKLILSNINQAGAGGLY					
		: : : : : : : : : : : : : : : : : :					
35	orf1-1	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGEILTSNINQAGAGGLY					
		360	370	380	390	400	410
		420	430	440	450	460	470
	orf1a.pep	FEGDFTVSPENNETWQAGAGVHISEDSTVTWKVNGVANDRLSKIGKGLTHVQAKGENQGS					
		:					
40	orf1-1	FQGDFTVSPENNETWQAGAGVHISEDSTVTWKVNGVANDRLSKIGKGLTHVQAKGENQGS					
		420	430	440	450	460	470
		480	490	500	510	520	530
	orf1a.pep	SVGDDGTVILDQQADDKGKKQAFSEIGLXSGRGTQVLNADNQFNPDKLYFGFRGGRLDLNG					
45	orf1-1	SVGDDGTVILDQQADDKGKKQAFSEIGLVSGRGTQVLNADNQFNPDKLYFGFRGGRLDLNG					
		480	490	500	510	520	530
		540	550	560	570	580	590
	orf1a.pep	HSLSFHRHQNTDEGAMIXXHNATTTSTVTTITGNESITQPSGKNINRLNYSKEIAYNGWFG					
50	orf1-1	HSLSFHRHQNTDEGAMIVNHNQDKESTVTTITGNKDIA-TGNN-NSLDSKKEIAYNGWFG					
		540	550	560	570	580	590

5	orf1a.pep	600 610 620 630 640 650 EKDTTKTNGRLNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLFSSGRPTPHAYNHLGSG
	orf1-1	600 610 620 630 640 650 EKDTTKTNGRLNLVYQPAEDRTLLLSGGTNLNGNITQTNGKLFSSGRPTPHAYNHLNDH
10	orf1a.pep	660 670 680 690 700 710 WSKMEGIPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKVEGDHLSNHAQAVFGV
	orf1-1	660 670 680 690 700 710 WSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNSHAQAVFGV
15	orf1a.pep	720 730 740 750 760 770 APHQSHTICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVXLXXXXXXXXLXGXAXLX
	orf1-1	720 730 740 750 760 770 APHQSHTICTRSDWTGLTNCVEKITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLN
20	orf1a.pep	780 790 800 810 820 830 GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNG
	orf1-1	780 790 800 810 820 830 GNLSANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNG
25	orf1a.pep	840 850 860 870 880 890 SLTSLDNAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSG
	orf1-1	840 850 860 870 880 890 SLTSLGNAKANVSHSALNGNVSLADKAVFHFESSRTGQISGGKDTALHLKDSEWTLPSG
30	orf1a.pep	900 910 920 930 940 TELGNLNLNDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVERFN
	orf1-1	900 910 920 930 940 950 TELGNLNLNDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVERFN
35	orf1a.pep	950 960 970 980 990 1000 TLTVNGKLNQGTFRFMSELFYRSDKLKLAESSEGTYYTLAVNNTGNEPVSLDQLTVVEG
	orf1-1	960 970 980 990 1000 1010 TLTVNGKLNQGTFRFMSELFYRSDKLKLAESSEGTYYTLAVNNTGNEPASLEQLTVVEG
40	orf1a.pep	1010 1020 1030 1040 1050 1060 KDNKPLSENLFNLQNEHVDAGAWRYQLIRKDGFRHLNPNVKEQELSDKLKGAEAKKQAE
	orf1-1	1020 1030 1040 1050 1060 1070 KDNKPLSENLFNLQNEHVDAGAWRYQLIRKDGFRHLNPNVKEQELSDKLKGAEAKKQAE
45	orf1a.pep	1070 1080 1090 1100 1110 1120 KDNAQSLDALIAAGRDAEKTESVAEPARXAGGENVGIMQAEKKRVQADKDSALAKQR
	orf1-1	1080 1090 1100 1110 1120 1130 KDNAQSLDALIAAGRDVAEKTESVAEPARQAGGENVGIMQAEKKRVQADKDTALAKQR
50	orf1a.pep	1130 1140 1150 1160 1170 1180 EAETRPXTTAFPRARXARRDLQPQPQPQPQPQORDLXSRYANSGLSEFSATLNSVFAV
	orf1-1	1140 1150 1160 1170 1180 1190 EAETRPATTAFPRARRARRDLQPQPQPQPQPQ--QRDLISRYANSGLSEFSATLNSVFAV

-475-

		1190	1200	1210	1220	1230	1240
	orf1a.pep	QDELDRVFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFS					
	orf1-1	QDELDRVFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFS					
5		1200	1210	1220	1230	1240	1250
	orf1a.pep	1250	1260	1270	1280	1290	1300
	orf1-1	HNRTENXFDDGIGNSARLAHGAVFGQYGIGRFDIGISTGAGFSSGXLSDGIGGKIRRRVL					
10		1260	1270	1280	1290	1300	1310
	orf1a.pep	1310	1320	1330	1340	1350	1360
	orf1-1	HNRTENTFDDGIGNSARLAHGAVFGQYIDRFYIGISAGAGFSSGSLSDGIGGKIRRRVL					
15		1320	1330	1340	1350	1360	1370
	orf1a.pep	1310	1320	1330	1340	1350	1360
	orf1-1	HYGIQARYRAGFGGFGIEPIYGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSF					
20		1370	1380	1390	1400	1410	1420
	orf1a.pep	KPAQHXSITPYXSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHA					
	orf1-1	KPAQHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHA					
25		1430	1440	1450			
	orf1a.pep	AAAKGPQLEAQHSAGIKLGYRWX					
	orf1-1	AAAKGPQLEAQHSAGIKLGYRWX					
		1440	1450				

Homology with adhesion and penetration protein hap precursor of *H.influenzae* (accession number P45387) (SEQ ID NO: 1153)

Amino acids 23-423 of ORF1 (SEQ ID NO: 648) show 59% aa identity with hap protein (SEQ ID NO: 1153) in 450aa overlap:

30	orf1	23	FXAAYLAICLSFGILPQAWAGHTYFGINYYQYRDF	82
	hap	6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYRDF	65
	orf1	83	KSMTKAPMIDFSVVS RNGVAALVGVQYIVSVAHNGGYNVDFGAEGXNIXDQXRXTYKIV	142
35	hap	66	TSMTKAPMIDFSVVS RNGVAALVGDQYIVSVAHNGGYNVDFGAEGN-PDQHRFTYQIV	124
	orf1	143	KRNYYKAGTKGHPYGGDYHMPRLHKKVTD AEPVEMTSYMDGRKYIDQNNYPDRVRIGAGR	202
	hap	125	KRNYYQAWERKHPYDGDYHMPRLHKKFVTEAEPVGMTTNDGKVYADRENYPERVRIGSGR	184
40	orf1	203	QYWRSDDEPNNRESSYHIA-----	222
	hap	185	QYWRD+D+DE N SSY+++	244
	orf1	223	-----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVF	277
	hap	245	SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF	304

orf1 278 --YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA 334
 Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A
 hap 305 QRYIPPINGHYSFVSNNNDGTGKLTLTRPSKDGSKAKSEVGTVKLFNPSLNQTAKEHV-KA 363

 orf1 335 AGGVNSYRPRLNNGENISFIDEKGKELILTSNINQGAGGLYFQGDFTV-SPENNETWQGA 393
 A G N Y+PR+ G+NI D+GKG L + +NINQGAGGLYF+G+F V +NN TWQGA
 hap 364 AAGYNIYQPRMEYGKNIYLGDOGKGTLTIENTNINQGAGGLYFEGNFVVGKQNNITWQGA 423

 orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGTL 423
 GV I +D+TV WKV+ NDRLSKIG GTL
 hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGTL 453

Amino acids 715-1011 of ORF1 (SEQ ID NO: 648) show 50% aa identity with hap protein (SEQ ID NO: 1153) in 258aa overlap:

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTLS 98
 DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS
 hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTILIDHSQFTLSNNATQTGNIKLS 792

 orf1 99 GNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158
 +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N
 hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTTVTLENATWTMPSDTTLQN 852

 orf1 159 LNLDNATITLNSAYRHDAAGAQTGSATDAPXXXXXXXXXXLLXVTPPTSVESRFNTLTVN 218
 L L+N+T+TLNSAY + S+ +AP L T PTS E RFNTLTVN
 hap 853 LTLNNSTVTLNSAY-----SASSNNAPRHRS-----LETETPTSAEHRFNTLTVN 899

 orf1 219 GKLNGQGTFRFMSELFYGRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKP 278
 GKL+GQGTFF+ S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP
 hap 900 GKLSGQGTFFQFTSSLFYKSDKLKLSNDAEGDYTLVRNTGKEPVTLEQLTLIESLDNKP 959

 orf1 279 LSENLNFTLQNEHVDAGA 296
 LS+ L FTL+N+HVDAGA
 hap 960 LSDKLKFTLENDHVDAGA 977

Amino acids 1192-1450 of ORF1 (SEQ ID NO: 648) show 41% aa identity with hap protein (SEQ ID NO: 1153) in 259aa overlap:

Orf1 1 LDRVFAEDRRNAVWTSIGIRDTKHYSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNR 60
 LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R
 hap 1135 LDRLFVDQAQSAVWTNIAQDKRRYDSDAFRAYQQKTNLQIGVQKALANGRIGAVFSSHR 1194

 orf1 61 TENTFDDGIGNSARLAHGA VFGQY GIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG 120
 ++NTFD+ + N A L + F QY K R+ ++YG
 hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGD LQFGVNVGTGISASKMAEEQSRKIHRKAINYG 1254

 orf1 121 IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA 180
 + A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P
 hap 1255 VNASYQFRLGQLGIQPYFGVNRYFIERENYQSEEV RVKTPSLAFNRYNAGIRVDYFTTPT 1314

 orf1 181 QHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA 240
 +IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +
 hap 1315 DNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKS 1374

orf1 241 KGPQLEAQHSAGIKLGYRW 259
 +G QL Q + G+KLG YRW
 hap 1375 QGSQLGKQNVGVKLG YRW 1393

Homology with a predicted ORF from *N.gonorrhoeae*

- 5 The blocks of ORF1 (SEQ ID NO: 648) show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) (SEQ ID NO: 654) from *N.gonorrhoeae*:

10	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN	60
	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDFGAEGXNIXDQXRXTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSY	179
	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSEDEPNNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGGNTFAQNGSG	239
20	orf1.pep	-----GSPMFIYDAQKQKWLINGVLQTNPNYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFGLPTGGSGSPMFIYDAQKQKWLINGVLQTNPNYIGKSNG	289
25	orf1.pep	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRT	359
	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQAGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDKGGELILTSNINQAGAGGLY	
30	orf1.pep	FQGDFTVSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQGAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
	orf1.pep	DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
35	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSDHA	803
	orf1ng	TFNGNL-VQAETRTIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
40	orf1.pep	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
	orf1ng	VQNGSLTSLSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWT	893

	orf1.pep	LPSGXELGNLNLNLDNATITLNSAYRHAAGAQTGSATDAPRRRSRRSRRSLLXVTPPTSVE	923
	orf1ng	LPSGTELGNLNLNLDNATITLNSAYRHAAGAQTGSAADAPRRRSRRS--LLSVTPPTSVE	950
5	orf1.pep	SRFNTLTVNGKLNQGGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLT	983
	orf1ng	SRFNTLTVNGKLNQGGTFRFMSELFGYRSGKLKLAESSEGTYTLAVNNTGNEPVSLEQLT	1010
10	orf1.pep	VVEGKDNKPLSENLFNFTLQNEHVDAGAW	1011
	orf1ng	VVEGKDNTPLSENLFNFTLQNEHVDAGAWRYQLIRKDGFEFRLHNPVKEQELSDKLGKAGET	1070
	orf1.pep	// LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
15	orf1.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFGQYGIGRFD	1299
	orf1.pep	IGISAGAGFSSGSLSDGIGKXRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1331
	orf1ng	IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1359
20	orf1.pep	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAASKVRTRVNTAVL	1391
	orf1ng	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAASKVRTRVNTAVL	1419
25	orf1.pep	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQSAGIKLGYRW	1440
	orf1ng	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQSAGIKLGYRW	1468

The complete length ORF1ng nucleotide sequence was identified [<SEQ ID 653>] (SEQ ID NO: 653):

30	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCTAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCCGCTTA	CTTAGCCATA	TGCCTGTCGT
	101	TCGGCATTCT	GCCCCAAGCC	CGGGCGGGAC	ACACTTATTT	CGGCATCAAC
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTTT	ACAACAAAAA	AGGGGAGTTG	GTCCGCAAAT
	251	CGATGACGAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTATC	GCGTAACGGC
35	301	GTGGCGGCAT	TGGCGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AATGTTGATT	TTGGTGCGGA	GGGAAGCAAT	CCCGATCAGC
	401	ACCGCTTTTC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	AGCAGGGACT
	451	AACGGCCATC	CTTATGGCGG	CGATTATCAT	ATGCCGCGTT	TGCACAAATT
40	501	TGTCACAGAT	GCAGAACCTG	TTGAGATGAC	CAGTTATATG	GATGGGTGGA
	551	AATACGCTGA	TTTAAATAAA	TACCCTGATC	GTGTTCGAAT	CGGAGCAGGC
	601	AGACAAATATT	GGCGGTCTGA	TGAAGACGAA	CCCAATAACC	GCGAAAGTTC
	651	ATATCATATT	GCAAGCGCAT	ATTCTTGGCT	CGTCGGTGGC	AATACCTTTG
	701	CACAAAATGG	ATCAGGTGGT	GGCACAGTCA	ACTTAGGTAG	CGAAAAAATT
	751	AAACATAGCC	CATATGGTTT	TTTACCAACA	GGAGGCTCAT	TTGGCGACAG
45	801	TGGCTACCA	ATGTTTATCT	ATGATGCCCA	AAAGCAAAAG	TGGTTAATTA
	851	ATGGGGTATT	GCAAACAGGC	AACCCCTATA	TAGGAAAAAG	CAATGGCTTC
	901	CAGCTAGTTC	GTAAAGATTG	GTTCTATGAT	GAAATCTTTG	CTGGAGATAC
	951	CCATTACGTA	TTCTACGAAC	CACATCAAAA	TGGGAAATAC	TTTTTTAACG
50	1001	ACAATAATAA	TGGCGCAGGA	AAAATCGATG	CCAAACATAA	ACACTATTCT
	1051	CTACCTTATA	GATTAATAAC	ACGAACCGTT	CAATTGTTTA	ATGTTTCTTT
	1101	ATCCGAGACA	GCAAGAGAAC	CTGTTTATCA	TGCTGCAGGT	GGGGTCAACA

5 1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCTT TATTGACAAA
1201 GGAAAAAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAAACGT
1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGATG GCAGTACCGT TACTTGGA
1351 GTAAACGGCG TGGCAAACGA CCGCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCTGGTTCAG GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGACG
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAAGCC
1501 TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC
10 1551 CGATAATCAG TTCAACCCCG ACAAACCTTA TTTCGGCTTT CGCGGCGGAC
1601 GTTTGGATTT GAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCAGCAAT AACACAACCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC
15 1851 GGATCGCACT TTACTGCTTT CCGGCGGAAC AAATTAAAC GGCAATATCA
1901 GCACAAACAAA CCGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGCACGCC
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCCACAAGG
2001 AGAAATCGTG TGGGACAACG ATTGGATCGA CCGCACATTT AAAGCGGAAA
2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCCGCAA TGTTGCCAAA
20 2101 GTGGAAGGCG ATTGGCATT T AAGCAATCAC GCCCAAGCAG TTTTCGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAGTTG TACCGAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCACGCTCA
2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGCAGGCG
25 2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCCAAA CGGCAACCTC
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAACCGG
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACACGCCG
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AAACGTAAGC
2551 CATTCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
30 2601 TTTTGAAAA AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCGGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCTATCG
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TGCGGCAGAT GCGCCGCGCC
2801 GCCGTTCGCG CCGTTCCTTA TTATCCGTTA CGCCGCCAAC TTCGGCAGAA
35 2851 TCCCCTTTCA ACACGCTGAC GGTAAACGGC AAATTGAACG GTCAGGGAAC
2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATGGAAGC
2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC
3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAAG GAAAAGACAA
40 3051 CACACCGCTG TCCGAAAATC TTAATTTCAC CCTGCAaaAc gaacacgtcg
3101 atgcccgcgc atggcGTTAT CAGCTTATCC gcaaaagacgG CGAGTTCCgc
3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGgc
3201 gggagaaACA GAgccgcctT TGACGGCAAA ACAGGCaaa CTTGCCGGca
3251 AAcaacagc ggaAAAAGAC AACgcgcaaa gccttgAcgc gctgattgcy
45 3301 gCcgggcgca atgccaccga AAAGGCAgaa agtgttgccg aaccgGCCCC
3351 GCAGGCAGGC GGGGAAAatg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA
3401 AACGGGTGCA GCGGATAAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG
3451 GAAACCCGGC CGGTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG
3501 GGATTTGCCG CAACCGCAGC CCCAACCGCA ACCCAACCG CAGCGCGACC
50 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC
3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA
3651 CCGCCGCAAC GCCGTTTGA CAAGCGGCAT CCGGGACACC AAACACTACC
3701 GTTCGCAAGA TTTCCGCGCC TACCGCAAAC AAACCGACCT GCGCCAAATC
3751 GGTATGCAGA AAAACCTCGG CAGCGGCGC GTCCGGATCC TGTTTTCGCA
55 3801 CAACCGGACC GGAAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC
3851 TTGCCACGG TGCCGTTTTT GGGCAATACG GCATCGGCAG GTTCGACATC
3901 GGCATCAGCG CGGGCGCGG TTTTAGTAGC GGCAGCCTT CAGACGGCAT
3951 CAGAGGCAAA ATCCGCCGCC GCGTGTGCA TTACGGCATT CAGGCAAGAT
4001 ACCGCGCAGG TTTCCGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG
4051 CGCTATTTTC TCCAAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC
60 4101 CACCCCGGGC CTTGCATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT
4151 ATTCATTCAA ACCGGCGCAA CACATTTCCA TCACGCCTTA TTTGAGCCTG

5

4201	TCCTATACCG	ATGCCGCTTC	CGGCAAAGTC	CGAACGCGCG	TCAATACCGC
4251	CGTATTGGCG	CAGGATTTCG	GCAAAACCCG	CAGTGC GGAA	TGGGGCGTAA
4301	ACGCCGAAAT	CAAAGGTTTC	ACGCTGTCCC	TCCACGCTGC	CGCCGCCAAG
4351	GGGCCGCAAT	TGGAAGCGCA	GCACAGCGCG	GGCATCAAAT	TAGGCTACCG
4401	CTGGTAA				

This is predicted to encode a protein having amino acid sequence [<SEQ ID 654>] (SEQ ID NO: 654):

10	1	MKTTDKRTE	THRKAAPTGR	IRFSPAYLAI	CLSGFILPQA	RAGHTYFGIN
	51	YQYYRDAFEN	KGKFAVGAKD	IEVYNKKGEL	VGKSMTKAPM	IDFSVVSRRNG
	101	VAALAGDQYI	VSAHNGGYN	NVDFGAEGSN	PDQHRFSYQI	VKRNRYKAGT
	151	NGHPYGGDYH	MPRLHKFVTD	AEPVEMTSYM	DGWKYADLNK	YPDRVRIGAG
	201	RQYWRSDEDE	PNNRESSYHI	ASAYSWLVG	NTFAQNGSGG	GTVNLGSEKI
15	251	KHSPYGLFPT	GGSGFDSGSP	MFYDAQKQK	WLTNGVLQGT	NPYIGKSNGE
	301	QLVRKDWFYD	EIFAGTHSV	FYEPHQNGKY	FFNDNNNGAG	KIDAKHKHYS
	351	LPYRLKTRTV	QLFNVSLSSET	AREPVYHAAG	GVNSYRPRLN	NGENISFIDK
	401	GKGELILTSN	INQGAGGLYF	EGNFTVSPKN	NETWQGAGVH	ISDGSTVTWK
	451	VNGVANDRLS	KIGKGTLLVQ	AKGENQGSVS	VGDGKVILDQ	QADDQGKKQA
20	501	FSEIGLVSGR	GTVQLNADNQ	FNPDKLYFGF	RGRRLDLNGH	SLSFHRIQNT
	551	DEGAMIVNHN	QDKESTVTIT	GKNDITTTGN	NNLDSKKEI	AYNGWFGEKD
	601	ATKTNGGLNL	NYPPEEADRT	LLSGGTNLN	GNITQTNGKL	FFSGRPTPHA
	651	YNHLGSGWSK	MEGIPQGEIV	WDNDWIDRTF	KAENFHIQGG	QAVVSRNVAK
	701	VEGDWHLNSH	AQAVFGVAPH	QSHITCTRSD	WTGLTSCTEK	TITDDKVIAS
25	751	LSKTDVRGNV	SLADHAHLNL	TGLATFNGNL	VQAETRTIRL	RANATQNGNL
	801	SLVGNAQATF	NQATLNGNTS	ASDNASFNL	NNAVQNGSLT	LSDNAKANVS
	851	HSALNGNVS	ADKAVFHFEN	SRFTGKISGG	KDTALHLKDS	EWTLPSGTEL
	901	GNLNLNATI	TLNSAYRHDA	AGAQTGSAAD	APRRRSRRSL	LSVTPPTSSE
	951	SRFNTLTVNG	KLNGQGTFRF	MSELPFYRSG	KLKLAESSEG	TYTLAVNNTG
30	1001	NEPVSLEQLT	VVEGKDNTP	SENLFNTLQN	EHVDAGAWRY	QLIRKDGEP
	1051	LHNPVKEQEL	SDKLKGAGET	EAALTAKQAG	LAQQAQAEK	NAQSLDALIA
	1101	AGRNATEKAE	SVAEPARQAG	GENAGIMQAE	EEKRVQADK	DTALAKQREA
	1151	ETRPATTAFF	RARRARRDL	PQPQPQPQP	QRDLISRYAN	SGLSEFSATL
	1201	NSVFAVQDEL	DRVFAEDRRN	AVWTSGIRDT	KHYRSQDFRA	YRQQTDLRQI
35	1251	GMQKNLGSGR	VGILFSHNRT	GNTFDDGIGN	SARLAHGAVF	GQYIGRFDI
	1301	GISAGAGFSS	GSLSDGIRGK	IRRRVLHYGI	QARYRAGFGG	FGTEPHIGAT
	1351	RYFVQKADYR	YENVNIATPG	LAFNRYRAGI	KADYSFKPAQ	HISITPYLSL
	1401	SYTDAASGKV	RTRVNTAVLA	QDFGKTRSAE	WGVNAEIKGF	TLSLHAAAAK
	1451	GPOLEAOHSA	GIKLGYRW*			

40 Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

ORF1-1 (SEQ ID NO: 650) and ORF1ng (SEQ ID NO: 654) show 93.7% identity in 1471 aa overlap:

		10	20	30	40	50	60
45	orf1-1.pep	MKTTDKR	TTETHRK	APKTGR	IRFSPAY	LAICLSF	GILPQAWAGHTYFGIN
	orf1ng-1	MKTTDKR	TTETHRK	APKTGR	IRFSPAY	LAICLSF	GILPQARAGHTYFGIN
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf1-1.pep	KGKFAVG	AKDIEV	YNKKG	ELVGS	MTKAP	MIDFSVVS

	orflng-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSVSRNGVAALAGDQYIVSVAHNGGYN	70	80	90	100	110	120
5	orfl-1.pep	NVDFGAEGRNPdqhrfTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSYM	130	140	150	160	170	180
	orflng-1	NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSYM	130	140	150	160	170	180
10	orfl-1.pep	DGRKYIDQNNYPDRVRIGAGRQYWRsDEEPNNRESSYHIASAYSswLVGGNTFAQNGSGG	190	200	210	220	230	240
	orflng-1	DGWKYADLNKYPDRVRIGAGRQYWRsDEEPNNRESSYHIASAYSswLVGGNTFAQNGSGG	190	200	210	220	230	240
15	orfl-1.pep	GTVNLGSEKIKHSPYGFLLPTGGsFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKsNGF	250	260	270	280	290	300
	orflng-1	GTVNLGSEKIKHSPYGFLLPTGGsFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKsNGF	250	260	270	280	290	300
20	orfl-1.pep	QLVRKDWfYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTV	310	320	330	340	350	360
	orflng-1	QLVRKDWfYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRTV	310	320	330	340	350	360
25	orfl-1.pep	QLFNvSLSETAREPVYHAAGGVNSYRPRLLNNGENISFIDEGKGELILTSNINQAGGLYF	370	380	390	400	410	420
	orflng-1	QLFNvSLSETAREPVYHAAGGVNSYRPRLLNNGENISFIDKGGKELILTSNINQAGGLYF	370	380	390	400	410	420
30	orfl-1.pep	QGDFTVSPENNETWQAGVHISEDSTVTWkVNGVANDRLSKIGKGTlHVQAKGENQGSIS	430	440	450	460	470	480
	orflng-1	EGNFTVSPKNNETWQAGVHISDGSTVTWkVNGVANDRLSKIGKGTLLVQAKGENQGSVS	430	440	450	460	470	480
35	orfl-1.pep	VGDGTvILDQQADDKGKKQAFSEIGLVSGRGTVQLNADNQFNPDkLYFGFRGGRLLDNGH	490	500	510	520	530	540
	orflng-1	VGDGKvILDQQADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPDkLYFGFRGGRLLDNGH	490	500	510	520	530	540
40	orfl-1.pep	SLSFHRIQNTDEGAMIVNHnQDKESTVTITGNKDIAATTGNnNSLDSKKEIAYNGWfGEKD	550	560	570	580	590	600
	orflng-1	SLSFHRIQNTDEGAMIVNHnQDKESTVTITGNKDITTTGNnNSLDSKKEIAYNGWfGEKD	550	560	570	580	590	600
45	orfl-1.pep	TTKTNGRLNLVYQPAEDRTLLLSGGTnLNgnITQTNGKLFFSGRPTPhAYNHLNDHWSQ	610	620	630	640	650	660
	orflng-1	ATKTNGRLNLNYQPEADRTLLLSGGTnLNgnITQTNGKLFFSGRPTPhAYNHLGSWSK	610	620	630	640	650	660

		670	680	690	700	710	720	
	orf1-1.pep	KEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNSHAQAVFGVAPH						
5	orf1ng-1	MEGIPOQGEIVWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLNSHAQAVFGVAPH	670	680	690	700	710	720
		730	740	750	760	770	780	
	orf1-1.pep	QSHTICTRSDWTGLTNCVEKTIITDDKVIASLTTKTDISGNVDLADHAHLNLTGLATLNGNL						
10	orf1ng-1	QSHTICTRSDWTGLTSCTEKTIITDDKVIASLSKTDIRGNVSLADHAHLNLTGLATLNGNL	730	740	750	760	770	780
		790	800	810	820	830	840	
	orf1-1.pep	SANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLT						
15	orf1ng-1	SAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNAVQNGSLT	790	800	810	820	830	840
		850	860	870	880	890	900	
	orf1-1.pep	LSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTEL						
20	orf1ng-1	LSDNAKANVSHSALNGNVSLADKAVFHFENS RFTGKISGGKDTALHLKDSEWTLPSGTEL	850	860	870	880	890	900
		910	920	930	940	950	960	
	orf1-1.pep	GNLNLNDNATITLNSAYRHDAAGAQTGSATDAPRRRRSRSSRLSVTPPTS SVESRFNTLT						
25	orf1ng-1	GNLNLNDNATITLNSAYRHDAAGAQTGSAADAPRRRSR---RSLSVTPPTS AESRFNTLT	910	920	930	940	950	
		970	980	990	1000	1010	1020	
	orf1-1.pep	VNGKLNQGQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDN						
30	orf1ng-1	VNGKLNQGQGTFRFMSELFGYRSGKLKLAESSEGTYTLAVNNTGNEPVSLEQLTVVEGKDN	960	970	980	990	1000	1010
		1030	1040	1050	1060	1070		
	orf1-1.pep	KPLSENLNFTLQNEHVDAGAWRYQLIRKDG EFR LHN P VKEQELSDKL GK A-----						
35	orf1ng-1	TPLSENLNFTLQNEHVDAGAWRYQLIRKDG EFR LHN P VKEQELSDKL GK AGETEAALTAK	1020	1030	1040	1050	1060	1070
		1080	1090	1100	1110	1120		
	orf1-1.pep	----EAKKQAEKDNAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAE E EKKRVQ						
40	orf1ng-1	QAQLAAKQAEKDNAQSLDALIAAGRNATEKAESVAEPARQAGGENAGIMQAE E EKKRVQ	1080	1090	1100	1110	1120	1130
		1130	1140	1150	1160	1170	1180	
	orf1-1.pep	ADKDTALAKQREAE TR PATTAFPRARRARRDL PQLPQPQPQPQRDLISRYANSGLSEFS						
45	orf1ng-1	ADKDTALAKQREAE TR PATTAFPRARRARRDL PQLPQPQPQPQPQRDLISRYANSGLSEFS	1140	1150	1160	1170	1180	1190
		1190	1200	1210	1220	1230	1240	
	orf1-1.pep	ATLNSVFVQDELDRVFAEDRRNAVWTS GIRD TKHYRSQDFRAYRQQTDLRQIGMQKNLG						
50	orf1ng-1	ATLNSVFVQDELDRVFAEDRRNAVWTS GIRD TKHYRSQDFRAYRQQTDLRQIGMQKNLG	1200	1210	1220	1230	1240	1250

-483-

5	orf1-1.pep	1250	1260	1270	1280	1290	1300
	orf1ng-1	SGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGI					
10	orf1-1.pep	1310	1320	1330	1340	1350	1360
	orf1ng-1	GGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR					
15	orf1-1.pep	1370	1380	1390	1400	1410	1420
	orf1ng-1	AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
20	orf1-1.pep	1430	1440	1450			
	orf1ng-1	KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX					

In addition, ORF1ng (SEQ ID NO: 654) shows 55.7% identity with hap protein (P45387) (SEQ ID NO: 1153) over a 1455aa overlap:

SCORES Init1: 1104 Initn: 4632 Opt: 2680
Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

25	orf1ng-1.pep	10	20	30	40	50	60
	p45387	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINQYYRDAEN					
30	orf1ng-1.pep	70	80	90	100	110	120
	p45387	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALAGDQYIVSAHNNGGYN					
35	orf1ng-1.pep	130	140	150	160	170	180
	p45387	NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSYM					
40	orf1ng-1.pep	190	200	210	220	230	240
	p45387	DGWKYADLNKYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG					
45	orf1ng-1.pep	250	260	270	280	290	300
	p45387	GTVNLGSEKIKHSPYGFPLPTGGSGFGSGSPMFIYDAQKQKWLINGVLQTGNPNYIGKSNFG					

	p45387	GYSYLGGDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQKWLINGILREGNPFEGKENG	220	230	240	250	260	270
5	orf1ng-1.pep	QLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRTV	310	320	330	340	350	360
	p45387	QLVRKSYF-DEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKS---GIPSEIK---I	280	290	300	310	320	
10	orf1ng-1.pep	QLFNVSLSETAREPVYHAA-GGVNSYRPRLNNGENISFIDKGKGELILTSNINQGAGGLY	370	380	390	400	410	419
	p45387	TLANMSLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQGSILIFASDINQGAGGLY	330	340	350	360	370	380
15	orf1ng-1.pep	FEGNFTVSPKNNETWQGAGVHISDGSTVTWVNGVANDRLSKIGKGTLLVQAKGENQGSV	420	430	440	450	460	479
	p45387	FEGNFTVSPNSNQTWQAGIHVSENSTVTWVNGVEHDRLSKIGKGTLLHVQAKGENKGS	390	400	410	420	430	440
20	orf1ng-1.pep	SVGDGKVILDQQADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGRLDLNG	480	490	500	510	520	539
	p45387	SVGDGKVILEQQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGRLDLNG	450	460	470	480	490	500
25	orf1ng-1.pep	HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITT-TGNN-NNLDSKKEIAYNGWFG	540	550	560	570	580	590
	p45387	HSLTFKRIQNTDEGAMIVNHNNTQAANVTITGNESIVLPNGNNINKLDYRKEIAYNGWFG	510	520	530	540	550	560
30	orf1ng-1.pep	EKDATKTNGRLNLNYQPEEADRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG	600	610	620	630	640	650
	p45387	ETDKNKHNGRLNLIYKPTTEDRTLLSGGTNLKGDITQTNGKLFFSGRPTPHAYNHLNKR	570	580	590	600	610	620
35	orf1ng-1.pep	WSKMEGIPQGEIWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLNHAQAVFGV	660	670	680	690	700	710
	p45387	WSEMEGIPQGEIWDHDWINRTFKAENFQIKGGSVVSRNVSSIEGNWTVSNANATFGV	630	640	650	660	670	680
40	orf1ng-1.pep	APHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDIRGNVSLADHAHLNLTGLATLN	720	730	740	750	760	770
	p45387	VPNQQNTICTRSDWTGLTTCQKVDLTDTKVINSIPKTQINGSINLTDNATANVKGLAKLN	690	700	710	720	730	740
45	orf1ng-1.pep	GNLSAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNAVQNG	780	790	800	810	820	830
	p45387	GNVTL-----TNHSQFTLSNNAQTQIG	750				760	770
	orf1ng-1.pep	SLTSLDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWTLPSG	840	850	860	870	880	890

p45387 NIRLSDNSTATVDNANLNGNVHLTDSAQFSLKNSHFSHQIQGDKGTVTLTLENATWTMPSD
780 790 800 810 820 830

5 orflng-1.pep 900 910 920 930 940 950
TELGNLNLDNATITLNSAYRHDAAQAQTGSAADAPRRRSRRSLLSVTPPTSASERSFNLT
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
p45387 TTLQNLTNNSTITLNSAY-----SASSNTPTRRRS---LETETPTSAEHRFNLT
840 850 860 870

10 orflng-1.pep 960 970 980 990 1000 1010
VNGKLNGQGTFRFMSELFYGRSGKLKLAESSEGYTLAVNNTGNPEVPVSLEQLTVVEGKDNL
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
p45387 VNGKLSGGTGFTSQSSLFYGYSKDLKLSNDAGDYILSVRNTEGPETLEQLTLVESKDN
880 890 900 910 920 930

15 orflng-1.pep 1020 1030 1040 1050 1060 1070
TPLSENLFNTLQNEHVDAWARYQLRKDGEFRLLHPNPKEQLSDKLKGAGETEAAALTAK
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
p45387 QPLSDKLKFTLENDHVDAGALRYKLVKNDEGFRLHNPIKEQLHNDLVRAEQARTLEAK
940 950 960 970 980 990

20 orflng-1.pep 1080 1090 1100 1110 1120 1130
QAQLAQQAEKDNAQSOLDALIAAGRNAS-EKAESVAEPARQAGGENAGIMQAEEEEKRV
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
p45387 QVEPTAKTTGEPKVRSRRAAFPDLPDQSLNLAELAQAE-LTAETQKSAAKTCKV
1000 1010 1020 1030 1040 1050

25 orflng-1.pep 1140 1150 1160 1170 1180 1190
QADK---DTALAKQREARETRPATTAFFPRARRARD-LPQPQPQPQPQRDLISRYANS
: : : : | : : : : | : : : : : : : : | : : : : : : : : : : : : : : : :
p45387 RSKRAVFSDPLLDSLFALLEALEVIDAPQQSEKDRLAQEBAEK-QKQKDLISRYNSA
1060 1070 1080 1090 1100 1110

30 orflng-1.pep 1200 1210 1220 1230 1240 1250
LSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTHKYRSQDFRAYRQQ-TDLRQIG
| | : | : | : | : | : | : | : : : | : : : | : : | : | : | : | : | : | : | :
p45387 LSELSTVNMSLSVQDELDRFLFDQAQSAWNIAQDKRRYDSADFAYQQQKTNLRQIG
1120 1130 1140 1150 1160 1170

35 orflng-1.pep 1260 1270 1280 1290 1300 1310
MQKNLGSGRGVGLFHSNRRTGNFTDDGIGNSARLAHGAVFGQYIGIFDIGISAGAGFSSG
: | : | : | : | : | : | : | : : | : : : | : : : | : : : | : : : | : : :
p45387 VQKALANGRIGAVFSHRSRDNTFDEQVKNHATLTMSGFAQYQWGDLQFQGVNVGTGISAS
1180 1190 1200 1210 1220 1230

40 orflng-1.pep 1320 1330 1340 1350 1360 1370
SLSDGIRGKIERRVLHYGIQARYRAGFGGIEPHIGATRYFVQKADRYENVNIATPGL
: : : | : | : : : | : : : | : | : : : | : | : : : | : | : : : | : | : : :
p45387 KMAEEQSRKIHKKAINYGYNASYQFRLGQLGIQPYFGVNYFIERENYQSEEVRVKPTPSL
1240 1250 1260 1270 1280 1290

45 orflng-1.pep 1380 1390 1400 1410 1420 1430
AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAASGKVRTVNATAVLAQDFGKTRSIEW
| | | | | : : | : : | : : | : : : : : : : : | : | : | : | : : | : | :
p45387 AFNRYNAGIRVDYTFPTDNISVKPFYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEV
1300 1310 1320 1330 1340 1350

1440 1450 1460 1469

```

orf1ng-1.pep  GVNAEIKGFTLSLHAAAAKGPQLEAQHSAGIKLGYRWX
                |::||| | :| : ::| || |:::|:|||||
p45387        GLKAEILHFQISAFISKSQGSQQLGKQQNVGVKLGYRW
                1360      1370      1380      1390

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 78

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 655] (SEQ ID NO: 655):

```

1  ..AAGGTGTGGC AATTTGTCTGA AGA.CCGCTG CGTGCCGTCG TGCCTGCCGA
51  CAGTTTTTGAA CCGACCGCGC AAAAATTGAA CCTGTTTAAG GCGGGTGCGG
101 CAACCATTTT GTTTTATGAA GATCAAAATG TCGTCAAAGG TTTGCAGGAG
151 CAGTTCCTTG CTTATGCCGC TAACTTCCCC GTTTGGGCGG ATCAGGCAAA
201 CGCGATGGTG CAGTATGCCG TTTGGACGAC ACTTGCCGCG GTCGGCGTAG
251 GTGCAAACCT GCAACATTAC AATCCCTTGC CCGATGCGGC GATTGCCAAA
301 GCGTGGAATA TCCCCGAAAA CTGGTTGTTG CGCGCACAAA TGGTTATCGG
351 CCGTATTGAA GGGGCGGCAG GTGAAAAGAC CTTTGAACCC GTTGCAGAAC
401 GTTTGAAAGT GTTCGGCGCA TAA

```

This corresponds to the amino acid sequence [SEQ ID 656; ORF6] (SEQ ID NO: 656; ORF6):

```

1  ..KVVQFVEXPL RAVVPADSFE PTAQKLNLFK AGAATILFYE DQNVVKGLQE
51  QFPAYAAANFP VWADQANAMV QYAVWTTLAA VGVGANLQHY NPLPDAAIAK
101 AWNIPENWLL RAQMVIIGIE GAAGEKTFEP VAERLKVFGA *

```

Further sequence analysis revealed a further partial DNA sequence [SEQ ID 657] (SEQ ID NO: 657):

```

1  ..CTGCGTGCCG TCGTGCCCTGC CGACAGTTTT GAACCGACCG CGCAAAAATT
51  GAACCTGTTT AAGGCGGGTG CGGCAACCAT TTTGTTTTAT GAAGATCAAA
101 ATGTCGTCAA AGGTTTGAGG GAGCAGTTCC CTGCTTATGC CGCTAACTTC
151 CCCGTTTGGG CGGATCAGGC AAACGCGATG GTGCAGTATG CCGTTTGGAC
201 GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT TACAATCCCT
251 TGCCCGATGC GGCGATTGCC AAAGCGTGGA ATATCCCCGA AAAGTGGTTG
301 TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG CAGGTGAAAA
351 GACCTTTGAA CCCGTTGCAG AACGTTTGAA AGTGTTTCGC GCATAA

```

This corresponds to the amino acid sequence [SEQ ID 658; ORF6-1] (SEQ ID NO: 658; ORF6-1):

```

1  ..LRAVVPADSF EPTAQKLNLF KAGAATILFY EDQNVVKGLQ EQFPAYAAANF
51  PVWADQANAM VQYAVWTTLA AVGVGANLQH YNPLPDAAIA KAWNIPENWL
101 LRAQMVIIGI EGAAGEKTFE PVAERLKVFG A*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF6 (SEQ ID NO: 656) shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) (SEQ ID NO: 660) from strain A of *N. meningitidis*:

```

5      orf6.pep      10      20      30
                        KVVWQFVEXPLRAVVPADSFEPTAQKLNLFK
orf6a      QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVWQFVEDALRAVVPADSFEPTAQKLNLFK
                        40      50      60      70      80      90

10     orf6.pep      40      50      60      70      80      90
                        AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
orf6a      AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
                        100     110     120     130     140     150

15     orf6.pep      100     110     120     130     140
                        NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
orf6a      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
                        160     170     180     190     200

```

20 The complete length ORF6a nucleotide sequence [SEQ ID 659] (SEQ ID NO: 659) is:

```

1  ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
51 TTCGTAAAT AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG
101 TCGAACACGC CGTTTTGCAC ACACCTTCTT CGTTCAATTC CCAATCTGCC
151 CGTGTGGTCG TGCTGTTTGG CGAAGAGCAT GATAAGGTGT GGCAATTTGT
25 201 CGAAGACGCG CTGCGTGCCG TCGTGCCTGC CGACAGTTT GAACCGACCG
251 CGCAAAATT GAACCTGTTT AAGGCGGGTG CGGCAACTAT TTTGTTTAT
301 GAAGATCAAA ATGTCGTCAA AGGTTTGCAG GAGCAGTTCC CTGCTTATGC
351 CGCCAACTTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG
401 CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAA CCTGCAACAT
30 451 TACAATCCCT TGCCCGATGC GCGCATTGCC AAAGCGTGA ATATCCCCGA
501 AAATGGTTG TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG
551 CAGGTGAAA GACCTTGAA CCAGTTGCAG AACGTTGAA AGTGTTCCGC
601 GCATAA

```

35 This is predicted to encode a protein having amino acid sequence [SEQ ID 660] (SEQ ID NO: 660):

```

1  MTRQSLQQA ESRRSIYSLN KNLPVGKDEI VQIVEHAVLH TPSSFNSQSA
51 RVVVLFGEED KVVWQFVEDA LRAVVPADSF EPTAQKLNLF KAGAAATILFY
101 EDQNVVKGLE EQFPAYANF PVWADQANAM VQYAVWTTLA AVGVGANLQH
151 YNPLPDAAIA KAWNIPENWL LRAQMVIIGI EGAAGEKTFE PVAERLKVFG
40 201 A*

```

ORF6a (SEQ ID NO: 660) and ORF6-1 (SEQ ID NO: 658) show 100.0% identity in 131 aa overlap:

-488-

		50	60	70	80	90	100
	orf6a.pep	TPSSFNSQSARVVVLFGEEDKQVQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY					
5	orf6-1						
					10	20	30
		110	120	130	140	150	160
	orf6a.pep	EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAAIA					
10	orf6-1						
		40	50	60	70	80	90
		170	180	190	200		
	orf6a.pep	KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFEPVAERLKVFGAX					
15	orf6-1						
		100	110	120	130		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 (SEQ ID NO: 656) shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) (SEQ ID NO: 662) from *N.gonorrhoeae*:

20	orf6.pep		KVWQFVEXPLRAVVPADSFEPTAQKLNLFK	30
	orf6ng	SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLKLFK		64
	orf6.pep	AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY		90
	orf6ng	AGAATILFYEDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGGAGANLQHY	:	124
25	orf6.pep	NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFEPVAERLKVFGA		140
	orf6ng	NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPVAERLKVFGA	:	174

The complete length ORF6ng nucleotide sequence [<SEQ ID 661>] (SEQ ID NO: 661) was identified as:

30	1	ATGGCCGTTG	CGTCAAATGT	CAGCTTGGAT	ATGTCCAATC	CTACGGTGTT
	51	ACGCATGGGA	TTACCCTTAT	ATATTGCGTC	CCTAAGAAGG	GGCGCAATAT
	101	ATAAGGTGTG	GCAATTGTG	GAAGACGCGC	TGCGTGCCGT	CGTGCCTGCC
	151	GACAGTTTGT	AACCGACCGC	GCAAAAATTG	AAGCTGTTTA	AGGCGGGCGC
	201	GGCAACCATT	TTGTTTTATG	AAGATCAAAA	TGTCGTCAA	GGTTTGCAGG
35	251	AGCAGTTCCC	TGCTTATGCC	GCCAACCTTC	CCGTTGGGC	GGACCAGGCG
	301	AACGCTATGG	TACAGTATGC	CGTCTGACG	ACACTTGCCG	CGGTCGGTGC
	351	AGGTGCAAAT	CTGCAACATT	ACAACCCCTT	GCCCGATGTG	GCGATTGCTA
	401	AAGCGTGGAA	TATTCCTGAA	AACTGGCTGT	TGCGCGCGCA	AATGGTTATC
	451	GGTGGTATTG	AAGGGGcggc	aggtgaaaaa	gtctttgaac	CCGTTGCgga
40	501	acgtttgAAA	GTGTTTCGGC	CATAA		

This encodes a protein having amino acid sequence [<SEQ ID 662>] (SEQ ID NO: 662):

1 MAVASNVSLD MSNPTVLRMG LPLYIASLRR GAIYKVWQFV EDALRAVVPA

51 DSFEPTAQKL KLFKAGAATI LFYEDQNVVK GLQEQFPAYA ANFPVWADQA
 101 NAMVQYAVWT TLAAVGAGAN LQHYNPLPDV AIAKAWNIPE NWLLRAQMVI
 151 GGIEGAAGEK VFEPVAERLK VFGA*

ORF6ng (SEQ ID NO: 662) and ORF6-1 (SEQ ID NO: 658) show 96.9% identity in 131 aa overlap:

					10	20	30
orf6-1.pep					LRAVVPADSFEPTAQKLNLFKAGAATILFY		
orf6ng	PTVLRMGLPLYIASLRRGAIYKVWFVEDALRAVVPADSFEPTAQKLKLFKAGAATILFY	20	30	40	50	60	70
		40	50	60	70	80	90
orf6-1.pep	EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA						
orf6ng	EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGAGANLQHYNPLPDVAIA	80	90	100	110	120	130
		100	110	120	130		
orf6-1.pep	KAWNIPENWLLRAQMVGIEGAAGEKTFEPVAERLKVFGAX						
orf6ng	KAWNIPENWLLRAQMVGIEGAAGEKVFEPVAERLKVFGAX	140	150	160	170		

It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 663>] (SEQ ID NO: 663)

```

1  ..GGCTACAAC  ACCTGTTGCG  GCGCGGCAGC  CGCATCGCCA  ACTACCAAAT
51  CAACGGCATC  CCCGTTGCCG  ACGCGCTGGC  CGATACGGG  CAATGCCAAC
101 ACCGCCGCCT  ATGAGCGCGT  AGAAGTCGTG  CGCGGCGTGG  CGGGGCTGCT
151 GGACGGCACG  GCGGAGCCTT  CCGCCACCGT  CAATCTGGTG  CGCAAACGCC
201 TGACCCGCAA  GCCATTGTTT  GAAGTCCGCG  CCGAAGCgGG  CAACCGcAAA
251 CATTTCGGGC  TGGACGCGGA  CGTATCGGGC  AGCCTGAACA  CCGAAG.crC
301 rCTGCGCgGC  CGCCTGGTTT  CCAcCTTCGG  ACGCGGCGAC  TCGTGGCGGC
351 GGCGCGAACG  CAGCCGskAT  GCCGAACCTT  ACGGCATTTT  GGAATACGAC
401 ATCGCACCGC  AAACCCGCGT  CCACGCAGrGC  ATGGACTACC  AGCAGGCGAA
451 AGAAACCGCC  GACGCGCCGC  TCAGcTACGC  CGTGTAACGAC  AGCCAAGGTT
501 ATGCCACCGC  CTTCGGCCCG  AAAGACAACC  CCGCCACAAA  TTGGGCGAAC
551 AGCCACCACC  GTGCGCTCAA  CCTGTTCGCC  GGCATCGAAC  ACCGCTTCAA
601 CCAAGACTGG  AAACCAAAG  CCGAATACGA  CTAC..

```

This corresponds to the amino acid sequence [<SEQ ID 664; ORF23>] (SEQ ID NO: 664; ORF23):

1 ..GYNVLFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL
 51 DGTGEPSTV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX
 101 LRGRVSTFG RGDWRRRER SRXAELYGIL EYDIAPQTRV HAXMDYQOAK
 151 ETADAPLSYA VYDSQGYATA FGPKNPATN WANSHHRLN LFAGIEHRFN
 5 201 QDWKLKAEYD Y..

Further work revealed the complete nucleotide sequence [<SEQ ID 665>] (SEQ ID NO: 665):

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA
 10 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC
 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
 251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCTGTGTGCA GGCGACCGGC
 301 ACCAGCCGCC AGATTTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT
 15 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG
 401 CGGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
 451 GTAGAAGTCG TGCGCGGCGT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCTGACCCGC AAGCCATTGT
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTTCG GCTGGACCGC
 20 601 GACGTATCGG GCAGCCTGAA CACCGAAGGC ACGCTGCGCG GCCGCCTGGT
 651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG
 701 ATGCCGAACCT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
 25 851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC
 901 AACCTGTTCG CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
 1051 GGTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCATTGAT
 30 1101 CGGCAAAATAC CGCCTGTTCG GCCGCGAACA CGATTTAATC GCGGGTATCA
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC
 1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
 1251 GCCTGCATCG TTTGCCCAAA CCATCCCGCA ATACGGCACC AGGCGGCAAA
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
 35 1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCTACACAG
 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC
 1501 AGCCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
 40 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG
 1601 AAGGCGGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
 1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT
 45 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC
 1951 ACGTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCGC CGAATCTGTC CTGAACGTGG ACAATCTGTT CAACAAACAC
 50 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence [<SEQ ID 666; ORF23-1>] (SEQ ID NO: 666; ORF23-1):

55 1 MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG

5
10
15
20
25
30
35
40
45
50
55
60
65
70

```

101  TSRIYGS DR AGNYLFARG SRIANYQING IPVADALADT GNANTAAYER
151  VEVVRGVAGL LDGTGEP SAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLDA
201  DVSGSLNTEG TLRGRLVSTF GRGDSWRRRE RSRDAELYGI LEYDIAPQTR
251  VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
301  NLFAGIEHRF NQDWKLKAEY DYTRSFRQRP YGVAGVLSID HNTAATDLIP
351  GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKGERSIIP
401  NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
451  ILGGRYTRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
501  SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNLNAS AAVYRARKNN
551  LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
601  DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHTDPA
651  TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH
701  YRTQPD RHY GALRTVNAAF TYRFK*

```

15 Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047) (SEQ ID NO: 1154)

ORF23 (SEQ ID NO: 664) and PupB protein (SEQ ID NO: 1154) show 32% aa identity in 205aa overlap:

20
25
30

```

Orf23  6  FARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRK 65
      ++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
PupB   215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273

Orf23  66  RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFXXXXXXXXXXXXXAE 125
      R T + + EAGN +G DVSG L +RGR V+ +
PupB   274 RPTAEAQASITGEAGNWDRYGTGFDVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL 333

Orf23  126 LYGILEYDIAPQTRVHAXMDYQQA KETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183
      +YGI E+D++ T + Y + D+PL + S G T N A +W+
PupB   334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRSLNAAPDWSY 391

Orf23  184 SHHRALNLFAGIEHRFNQDWKLKAE 208
      + H + F IE + W K E
PupB   392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 (SEQ ID NO: 664) shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) (SEQ ID NO: 668) from strain A of *N. meningitidis*:

35
40

```

orf23.pep          10      20      30
                  GNYNLFARGSRIANYQINGIPVADALADTG
                  |||||
orf23a.            QMRDQNIKALDRALLQATGTSRQIYGS DRAGNYNLFARGSRIANYQINGIPVADALADTG
                  90      100     110     120     130     140

                  40      50      60      70      80      90
orf23.pep          NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD

```

	orf23a	NANTAAAYERVEVVRGVAGLLDGTGEPSATVNLRVKRPTRKPLFEVRAEAGNRKHFGLGAD 150 160 170 180 190 200
5	orf23.pep	VSGSLNTEXXLRGRLVSTFGRGDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAK 100 110 120 130 140 150
	orf23a	VSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGILEYDIAPQTRVHAGMDYQQAK 210 220 230 240 250 260
10	orf23.pep	ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNFLFAGIEHRFNQDWKLKAeyd 160 170 180 190 200 210
	orf23a	ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHREALNFLFAGIEHRFNQDWKLKAeyd 270 280 290 300 310 320
15	orf23.pep	Y
	orf23a	YTRSFRFPYGVAGVLSIDHNNTAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIA 330 340 350 360 370 380

20 The complete length ORF23a nucleotide sequence [<SEQ ID 667>] (SEQ ID NO: 667) is:

25	1	ATGACACGCT	TCAAAATATTC	CCTGCTGTTT	GCCGCCCTGT	TGCCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCAAAACCG	CAGGAAAGCA
	101	CTGAATTGCC	GACCATCACC	GTTACGCGCG	ACCGCACCGC	GAGTTCCAAC
	151	GACGGCTACA	CTGTTTCCGG	CACGCACACC	CCGCTCGGGC	TGCCCATGAC
	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
30	251	CGCACC AAAA	CATCAAAGCG	CTCGACCGCG	CCCTTTGCGA	GGCAGCCGGC
	301	ACCAGCCGCG	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACTTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACTACCA	AATCAACGGC	ATCCCCGTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGCGCT	GGCGGGCGTG	CTGGACGGCA	CGGGCGAGCC
35	501	TTCCGCCACC	GTCAATCTGG	TGCGCAAACG	CCC GACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCG	GGCAACCGCA	AACATTTGCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	TGCCGAAGGC	ACGCTGCGCG	GCCCGCTGGT
	651	TTCACCTTTC	GGACGCGGCG	ACTCGTGGCG	ACGACGCGAA	CGCAGCCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCCGC
40	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CCGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CGAAAGACAA	CCCCGCCACA	AATTGGGCGA	ACAGCCGCCA	CCGTGCGCTC
	901	AACCTGTTTC	CCGGCATCGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GCAGCCGCTT	CCGCCAGCCC	TAGCGCGTAG
45	1001	CAGGCGTGCT	TTCCATCGAC	CACAACACCG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCCGACCC	GCGCACCCAC	AGCGCCAGCG	TGTCATTAAT
	1101	CGGCAAAATAC	CGCCTGTTTC	GCCGCGAACA	CGATTTAATC	GCGGGTATCA
	1151	ACGGTTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATCCCC
	1201	AACGCCATTTC	CCAACGCCTA	CGAATTTTCC	CGCACGGGTG	CCTACCCGCA
50	1251	GCCTGCATCG	TTTGCCCCAAA	CCATCCCGCA	ATACGGGCACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGTGT
	1351	ATACTCGGCG	GCAGATACAG	CCGTTACCGC	ACCGGCAGCT	ACGACAGCCG
	1401	CACACAAGGC	ATGACCTATG	TGTCCGCCAA	CCGTTTCACC	CCCTACACAG
	1451	GCATCGTGTT	CGACCTGACC	GGCAACCTGT	CGCTTTACGG	CTCGTACAGC
55	1501	AGCCTGTTTC	TCCCGCAATC	GCAAAAAGAC	GAACACGGCA	GCTACCTGAA
	1551	ACCCGTAACC	GGCAACAATC	TGGAAGCCGG	CATCAAAGGC	GAATGGCTTG
	1601	AAGGCCGTCT	GAACGCATCC	GCCGCCGTGT	ACCGCGCCCG	TAAAAACAAC
	1651	CTCGCCACCG	CAGCAGGACG	CGACCCGAGT	GGCAACACCT	ACTACCGCGC
	1701	CGCCAACCAA	GCCAAAACCC	ACGGCTGGGA	AATCGAAGTC	GGCGGCCGCA

5
10
15
20
25

```

1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTACCCG AACGCAGCTT
1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC
1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCGC CGAACTGTCG CTGAACGTGG ACAATCTGTT CAACAAACAC
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACCTG GGACAGTGAA
2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

```

This encodes a protein having amino acid sequence [SEQ ID 668] (SEQ ID NO: 668):

15
20
25

```

1  MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN
51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKA LDRALLQATG
101 TSRQIYGSDR AGYNLYFARG SRIANYQING IPVADALADT GNANTAAYER
151 VEVVRGVAGL LDGTGEPSAT VNLVRKRPTR KPLFEVRAEA GNRKHFGLGA
201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQP YGVAGVLSID HNTAATDLIP
351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
451 ILGGYRSRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNLAS AAVYRARKNN
551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHTDPA
651 TLRIPNPAAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVDNLFNKH
701 YRTQPDHRSY GALRTVNAAF TYRFK*

```

ORF23a (SEQ ID NO: 668) and ORF23-1 (SEQ ID NO: 666) show 99.2% identity in 725 aa overlap:

30
35
40
45
50

```

          10      20      30      40      50      60
orf23a.pep MTRFKYSLLF AALLPVYAQADVSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
          |||
orf23-1    MTRFKYSLLF AALLPVYAQADVSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
          10      20      30      40      50      60

          70      80      90      100     110     120
orf23a.pep PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNLYFARG
          |||
orf23-1    PLGLPMTLREIPQSVSVITSQQMRDQNIKTLD RALLQATGTSRQIYGSDRAGYNLYFARG
          70      80      90      100     110     120

          130     140     150     160     170     180
orf23a.pep SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRPTR
          |||
orf23-1    SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRLTR
          130     140     150     160     170     180

          190     200     210     220     230     240
orf23a.pep KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRQRRSRDAELYGI
          |||
orf23-1    KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRLVSTFGRGDSWRRRRSRDAELYGI
          190     200     210     220     230     240

          250     260     270     280     290     300
orf23a.pep LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL

```

-494-

	orf23-1	 LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL 250 260 270 280 290 300
5	orf23a.pep	310 320 330 340 350 360 NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH
	orf23-1	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH 310 320 330 340 350 360
10	orf23a.pep	370 380 390 400 410 420 SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS
	orf23-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS 370 380 390 400 410 420
15	orf23a.pep	430 440 450 460 470 480 FAQTIPQYGTRRQIGGYLATRFRAADNLSLILGGRYSRYRTGSYDSRTQGMTYVVSANRFT
	orf23-1	FAQTIPQYGTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVVSANRFT 430 440 450 460 470 480
20	orf23a.pep	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLNAS
	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLNAS 490 500 510 520 530 540
25	orf23a.pep	550 560 570 580 590 600 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR
	orf23-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR 550 560 570 580 590 600
30	orf23a.pep	610 620 630 640 650 660 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAAK
	orf23-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAAK 610 620 630 640 650 660
35	orf23a.pep	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFPRAELSLNVDNLFNKHRYTQPDRHSYGALRTVNAAF
	orf23-1	ARAADNSRQKAYAVADIMARYRFPRAELSLNVDNLFNKHRYTQPDRHSYGALRTVNAAF 670 680 690 700 710 720
40	orf23a.pep	TYRFXK
	orf23-1	TYRFXK

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 (SEQ ID NO: 664) shows 93.4% identity over a 211aa overlap with a predicted ORF

45 (ORF23.ng) (SEQ ID NO: 670) from *N. gonorrhoeae*:

	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPD	60
5	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTRKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGR	120
	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAF	180
10	orf23.pep	GPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNSRNRALNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence [<SEQ ID 669>] (SEQ ID NO: 669) is predicted to encode a
 15 protein comprising amino acid sequence [<SEQ ID 670>] (SEQ ID NO: 670):

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTRKP	LFEVRAEAGN	RKHFGGLADV
20	101	SGSLNAEGTL	RGRLVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPQTRVH
	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSRNRALNL
	201	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
25	301	IPNAYEFSRT	GAYPQPSSFA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGRYSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
	401	FVPQLQKDEH	GSYLKPVTGN	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
	451	TAAGRDQSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQ	QAGYSQSKPRDQ
	501	DGSRLNPD	SV PERSFKLFTA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTELSLN	VDNLFNKHRYR
	601	TQPD	RHSYGA LRTVNAAFTY	RFK*		

30 Further work revealed the complete nucleotide sequence [<SEQ ID 671>] (SEQ ID NO: 671):

	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
35	101	CCGAATTGCC	GACCTACACC	GTTACCGCGC	ACCGCACCGC	GAGTTCACAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCATGAC
	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
40	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAATAACCA	AATCAACGGC	ATCCCGTGTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGGACGGCA	CGGGCGAGCC
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCTCGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
50	701	ATGCCGAAC	TACGGCATT	TTGGAATACG	ACATCGCACC	GCAAAACCGC
	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCTGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCcgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC

5
10
15
20

```

1101 CGGCAAATAC CgcctGTTTCG GCCGCGAGCA CGATTTTAATC GCGGGTATCA
1151 ACGGCTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATTCCC
1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCACGGGCG CCTATCCGCA
1251 GCCATCATCG TTTGCCCAAA CCATCCCGCA ATACGACACC AGGCGGCAAA
1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
1351 ATACTCGGCG GCAGATACAG CCGCTACCGC GCAGGCAGCT ACAACAGCCG
1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG
1451 GCATCGTGTT CGATCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC
1501 AGCCTGTTTC TCCGCAATT GCAAAAAGAC GAACACGGCA GCTACCTGAA
1551 ACCCGTAACC GGCAACAATC TGGAAGCCGA CATCAAAGGC GAATGGCTTG
1601 AAGGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
1651 CTCGCCACCG CAGCAGGACG CGACCAGAGC GGCAACACCT ACTATCGCGC
1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCCGA
1751 TCACGCCCCA ATGGCAGATA CAGGCAGGCT ACAGCCAAAG CAAACCCCGC
1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTAcCCG AACGCAGCTT
1851 CAAACTCTTC ACCGCCTACC ACTTAGCCCC CGAAGCCCCC AGCGGCCGGA
1901 CCATcggtGC GGGTGTGCGC CGGCAGGGCG AAACCCACAC CGACCCAGCC
1951 GCGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG TCGCCAACAG
2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCAC CGAACTGTCT CTGAACGTGG ACAACCTGTT CAACAAACAC
2101 TACCGCACCC AGCCCCACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
2151 CGCGGCGTTT ACCTATCGGT TTAATAAA

```

This corresponds to the amino acid sequence [<SEQ ID 672; ORF23ng-1>] (SEQ ID NO: 672;

25 ORF23ng-1):

30
35
40

```

1 MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN
51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
101 TSRQIYGSDR AGYNILFARG SRIANYQING IPVADALADT GNANTAAYER
151 VEVVRGVAGL PDGTGEP SAT VNLVRKHPT R KPLFEVRAEA GNRKHFGLGA
201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSRNRAL
301 NLFAGIEHRF NQDWKLKAEY DYTRSFRFP YGVAGVLSID HSTAATDLIP
351 GYWHADPRTH SASMSLTGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401 NAIPNAYEFS RTGAYPQPSS FAQTIPQYDT RRQIGGYLAT RFRAADNLSL
451 ILGGYRSRYR AGSYNSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
501 SLFVPQLQKD EHGSYLKPV T GNNLEADIKG EWLEGRNLNAS AAVYRARKNN
551 LATAAGRDQS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKPR
601 DQDGSRLNPD SVPERSFKLF TAYHLAPEAP SGRITIGAGVR RQGETHTDPA
651 ALRIPNPAAK ARAVANSRQK AYAVADIMAR YRFNPRTELS LNVDNLFNKH
701 YRTQPDHRSY GALRTVNAAF TYRFK*

```

ORF23ng-1 (SEQ ID NO: 672) and ORF23-1 (SEQ ID NO: 666) show 95.9% identity in 725 aa overlap:

45
50

```

          10      20      30      40      50      60
orf23-1.pep MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT
          |||
orf23ng-1   MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT
          10      20      30      40      50      60

          70      80      90     100     110     120
orf23-1.pep PLGLPMTLREIPQSVSVITSQQMRDQNIKT LDRALLQATGTSRQIYGSDRAGYNILFARG
          |: |||
orf23ng-1   PFGLPMTLREIPQSVSVITSQQMRDQNIKT LDRALLQATGTSRQIYGSDRAGYNILFARG
          70      80      90     100     110     120

```


-497-

5	orf23-1.pep	130 140 150 160 170 180	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTR
	orf23ng-1	130 140 150 160 170 180	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPDGTGEP SATVNLVRKHPTTR
10	orf23-1.pep	190 200 210 220 230 240	KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRLVSTFGRGDSWRRRERSRDAELYGI
	orf23ng-1	190 200 210 220 230 240	KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRLERSRDAELYGI
15	orf23-1.pep	250 260 270 280 290 300	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL
	orf23ng-1	250 260 270 280 290 300	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRNRAL
20	orf23-1.pep	310 320 330 340 350 360	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHN TAATDLIPGYWHADPRTH
	orf23ng-1	310 320 330 340 350 360	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHSTAATDLIPGYWHADPRTH
25	orf23-1.pep	370 380 390 400 410 420	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYQPAS
	orf23ng-1	370 380 390 400 410 420	SASMSLTGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYQPSS
30	orf23-1.pep	430 440 450 460 470 480	FAQTIPQYGTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMYVSANRFT
	orf23ng-1	430 440 450 460 470 480	FAQTIPQYDTRRQIGGYLATRFRAADNLSLILGGRYSTRYRAGSYNSRTQGMYVSANRFT
35	orf23-1.pep	490 500 510 520 530 540	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS
	orf23ng-1	490 500 510 520 530 540	PYTGIVFDLTGNLSLYGSYSSLFVPLQKDEHGSYLKPVTGNNLEADIKGEWLEGRNLAS
40	orf23-1.pep	550 560 570 580 590 600	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR
	orf23ng-1	550 560 570 580 590 600	AAVYRARKNNLATAAGRDQSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKPR
45	orf23-1.pep	610 620 630 640 650 660	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSETHDPA TLRI PNPAAK
	orf23ng-1	610 620 630 640 650 660	DQDGSRLNPDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRI PNPAAK
50	orf23-1.pep	670 680 690 700 710 720	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPD RHSYGALRTVNAAF
	orf23ng-1	670 680 690 700 710 720	ARAVANSRQKAYAVADIMARYRFNPRTSLNVDNLFNKHYRTQPD RHSYGALRTVNAAF

orf23-1.pep TYRFKX
 |||||
orf23ng-1 TYRFKX

5

In addition, ORF23ng-1 (SEQ ID NO: 672) shows significant homology with an OMP (SEQ ID NO: 1155) from *E.coli*:

10

sp|P16869|FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-
FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR)gi|1651542|gnl|PID|d1015403
(D90745) Outer membrane protein FhuE precursor [Escherichia coli]
)gi|1651545|gnl|PID|d1015405 (D90746) Outer membrane protein FhuE precursor
[Escherichia coli])gi|1787344 (AE000210) outer-membrane receptor for Fe(III)-
coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor
[Escherichia coli] Length = 729

15

Score = 332 bits (843), Expect = 3e-90
Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)

Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRL 95
T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL +
Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSTAGTKMQMTQRDIPQSVTIVSQQRMEDQQLQTLGEVM 102

20

Query: 96 LQATGTSRQIYGSDRAGYNLFFARGSRANYQINGIP-----VADALADTGNANTAA 147
G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A
Sbjct: 103 ENTLGISKSDALY---YSRGFQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154

25

Query: 148 YERVEVVRGVAGLPDGTGEPSTVNLVRKHPTKPLF-EVRAEAGNRKHFGLGADVSGSL 206
+ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L
Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSP 214

Query: 207 NAEGLRGLVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQAKETADA 266
+G +R R+V + DSW S GI++ D+ T + AG +YQ+ +
Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDYRNSKTFFSGIVDADLGLTTLTSLAGYEQRIDVNSPT 274

30

Query: 267 PLSYAVYDSQGYATAFGPKDNPNATNWSNSNRNALNLFAGIEHRFNQDWKLKAEYDYTRSR 326
+++ G + ++ + A +W+ + +F ++ +F W+ ++
Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334

Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLFG 374
F + Y A V D ++ PG+ W++ R A + G Y LFG
Sbjct: 335 FDSKMMYVDAYVKNADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394

35

Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432
R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
Sbjct: 395 RQHNLMMFG-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451

40

Query: 433 QIGGYLATRFRAADNLSLILGGYRSRYRAGSYNSRTQGPTY-VSANRFTPYTGIVFDXXX 491
Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504

Query: 492 XXXXXXXXXXXXFPQLQKDEHGSYLKPVGTGNNLEADIKGEWLEGRNLASAAVYRARKNNL 551
F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
Sbjct: 505 NWSTYASYTSIFQPQNDRDSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564

45

Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608
A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624

Query: 609 PDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAAKARAVANSR 668
 P ++P + K+FT+Y L P P T+G GV Q +TD P RA
 Sbjct: 625 P-NLPRTTVKMFTSYRL-PVMPE-LTVGGGVNWQNRVYTDTV-----TPYGTFR-----E 672

Query: 669 QKAYAVADIMARYRFNPRTELSLNDNLFNKHYRTQPDRH-SYGALRTVNAAFTYRF 724
 Q +YA+ D+ RY+ L NV+NLf+K Y T + YG R. + TY+F
 Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTPrNFSITGTYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (SEQ ID NO: 666) (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 (SEQ ID NO: 666) is a surface-exposed protein, and that it is a useful immunogen.

Example 80

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 673>] (SEQ ID NO: 673):

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
 101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTC
 151 AGCGTCAGcA CGCCTGCTTC GGCGgcGgCa ATCATACCTT CGTCTTCGGA
 201 AACGGGGATA AACGcGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
 251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
 25 301 CCGTGCCTAC CGCAGACGCT CAAGCCCAT TnTTCAAGAA TGCGTGCCAC
 351 TnAGTCGCCG ACGGGG..

This corresponds to the amino acid sequence [<SEQ ID 674; ORF24>] (SEQ ID NO: 674; ORF24):

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
 101 PCVPQTLKPI XSRMRATXSP TG..

Further work revealed the complete nucleotide sequence [<SEQ ID 675>] (SEQ ID NO: 675):

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA

-500-

5
10
15

```

101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
151 AGCGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
201 AACGGGGATA AACGCGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA
251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCCTAC CGCAGACGCT CAAGCCCATT TCTTCAAGAA TGCGTGCCAC
351 TGAGTCGCCG ACGGCGGGGG TCGGCGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCCACG
451 CGGGTAATTT TGAAAGCAGT TTTCTTCACT ACTTCCGCAA CTTGCGGTCAA
501 TGTCGTTGCA TCTGAATTTT CCAACGCGGC TTTTACGACA CCTGGGCGCG
551 ATACGCGGAC ATTGATAACG GCATCCGCTT CGCCCGAACC ATGAAACGCG
601 CCCGCCATAA ACGGGTTGTC TTCCACCGCG TTGCAGAACA CGACAATTTT
651 AGCGCAGCCG AAACCTTCGG GCGTGATTTC CGCCGTGCGT TTGACGGTTT
701 CGCCCGCCAG CTTGACCGCA TCCATATTGA TACCGGCACG CGTACTGCCG
751 ATATTGATGG AGCTGCACAC AATATCGGTA GTCTTCATCG CTTGCGGAAT
801 GGAGCGGATT AACACCTCAT CCGAAGGCGA CATCCCTTTT TGCACCAACG
851 CGGAAAAACC GCCGATAAAA GACACACCGA TGGCTTTGGC AGCTTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA

```

This corresponds to the amino acid sequence [<SEQ ID 676; ORF24-1>] (SEQ ID NO: 676;

20 ORF24-1):

25

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSST
151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDPTPLIT ASASPEP*NA
201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LTVSPASLTA SILIPARVLP
251 ILMELHTISV VFIASGMERI NTSSEGDIPF CTNAEKPPIK DTPMALAALS
301 KVCATLT*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 (SEQ ID NO: 674) shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) (SEQ ID NO: 678) from strain A of *N. meningitidis*:

35

```

      10      20      30      40      50      60
orf24a.pep MRTAVVLLLIIMPMAASSAMMPEMVCAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
          |||
orf24      MRTAVVLLLIIMPMAASSAMMPEMVCAGVSPGTAIISKPTEQTAVMASSLSNVSTPASAAA
          |||
      10      20      30      40      50      60

      70      80      90     100     110     120
orf24a.pep IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVPCVPQTLKPISSRMRESPT
          |||
orf24      IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVPCVPQTLKPISSRMRESPT
          |||
      70      80      90     100     110     120

      130     140     150     160     170     180
orf24a.pep TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
          |||
orf24      TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
          |||
      130     140     150     160     170     180

```

45

-501-

		190	200	210	220	230	240
	orf24a.pep	PGPDTP	LITASASPEPXNAP	AI	XLQNTTILAQPKPSSVISXVRL	LMVSPASLTA	
5	orf24	PGPDTP	LITASASPEPXNAP	INGLSSTALQNTTILAQPKPSG	VISAVRLTVSPASLTA		
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf24a.pep	SILIPARVLPILMELHTISVVF	FIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS				
10	orf24	SILIPARVLPILMELHTISVVF	FIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS				
		250	260	270	280	290	300
	orf24a.pep	KVCATLTX					
15	orf24	KVCATLTX					

The complete length ORF24a nucleotide sequence [[SEQ ID 677](#)] ([SEQ ID NO: 677](#)) is:

	1	ATGCGCACGG	CAGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
	51	GGCAATGATG	CCGGAAATGG	TGTGCGCGGG	TGTGTCGCCG	GGAACGGCAA
20	101	TCATATCCAA	NCCGACCGAA	CAAACGGCGG	TCATCGCTTC	GAGTTTATCC
	151	AACGTCAGCA	CGCCTGCTTC	GGCGGCGGCA	ATCATACCTT	CGTCTTCGGA
	201	NACGGGGATA	AACGCGCCAC	TCAAACCGCC	AACCGCGCTC	GAAGCCATCA
	251	TGCCGCCCTT	TTTCACGGCA	TCGTTTCAGCA	ATGCCAAAGC	TGCTGTGTG
	301	CCGTGCGTAC	CGCAGACGCT	CAAACCCATT	TCTTCAAGAA	TGCGCGCCAC
	351	CGAGTCGCCG	ACGGCAGGGG	TCGGTGCCAG	CGACAAGTCG	AGAATACCAA
25	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GGCCGATGAG	TTCCGCCACG
	451	CGGGTAATTT	TGAAGGCGGT	TTTCTTCACA	ACTTCGGCAA	CTTCGGTCAA
	501	TGTCGTTGCA	TCCGAATTTT	CCAACGCGGC	TTTACGACA	CCCGGGCCGG
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCTGAGCC	GTGAAACGCG
	601	CCCGCCATAN	ACGGGTTGTC	TTCCNCCGCG	TTGCAGAAAC	CGACGATTTT
30	651	GGCGCAGCCG	AAACCTTCTA	GTGTGATTTC	ANCCGTGCGT	TTGATGGTTT
	701	CGCCCGCCAG	TCTGACCGCG	TCCATATTGA	TACCGGCGCG	CGTACTGCCG
	751	ATATTGATGG	AGCTGCACAC	GATATCAGTA	GTCTTCATCG	CTTCGGGAAT
	801	GGAACGGATN	AACACCTCGT	CAGAAGGCGA	CATACCTTTT	TGCACCAGCG
	851	CGGAAAAGCC	GCCAATAAAA	GACACGCCGA	TGGCTTTGGC	AGCCTTATCC
35	901	AAAGTTTGCG	CCACGCTGAC	GTAA		

This encodes a protein having amino acid sequence [[SEQ ID 678](#)] ([SEQ ID NO: 678](#)):

	1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIISXPTE	QTAVIASSLS
	51	NVSTPASAAA	IIPSSSXTGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAVV
40	101	PCVPQTLKPI	SSRMATESP	TAGVGASDKS	RIPNGIFSIF	EASRPMSSPT
	151	RVILKAVFFT	TSATSVNVVA	SEFSNAAFTT	PGPDTPTLIT	ASASPEP*NA
	201	PAIXGLSSXA	LQNTTILAQP	KPSSVISXVR	LMVSPASLTA	SILIPARVLP
	251	ILMELHTISV	VFIASGMERX	NTSSEGDIPF	CTSAEKPPIK	DTPMALAALS
45	301	KVCATLT*				

It should be noted that this protein includes a stop codon at position 198.

ORF24a ([SEQ ID NO: 678](#)) and ORF24-1 ([SEQ ID NO: 676](#)) show 96.4% identity in 307 aa overlap:

-502-

		10	20	30	40	50	60
	orf24a.pep	MRTAVVLLLLIMPMAASSAMPEMVCAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA					
5	orf24-1	MRTAVVLLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTQTAVMASSLSNVSTPASAAA					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf24a.pep	IIPSSSXTGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPQTLKPISSRMATESP					
10	orf24-1	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPQTLKPISSRMATESP					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf24a.pep	TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAFFT					
15	orf24-1	TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAFFT					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf24a.pep	PGPDTPTLITASASPEPXNAPAI XGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA					
20	orf24-1	PGPDTPTLITASASPEPXNAPAI NGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf24a.pep	SILIPARVLPILMELHTISVVF IASGMRXNTSSEGDIPFCTSAEKPPIKDTPMALAALS					
25	orf24-1	SILIPARVLPILMELHTISVVF IASGMRINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
	orf24a.pep	KVCATLTX					
	orf24-1	KVCATLTX					

30 Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 (SEQ ID NO: 674) shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) (SEQ ID NO: 680) from *N.gonorrhoeae*:

	orf24.pep	MRTAVVLLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTQTAVMASSLSNVSTPASAAA	60
35	orf24ng	MRTAVVLLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTQTAVMASSLSNVSTPASAAA	60
	orf24.pep	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPQTLKPIXSRMRATXSP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPQTLKPISSRMATESP	120
	orf24.pep	TG	122
40	orf24ng	TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVRLTASEFSNAALTT	180

The complete length ORF24ng nucleotide sequence [<SEQ ID 679>] (SEQ ID NO: 679) is:

1 ATGCGCACGG CGGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC

20 This encodes a protein having amino acid sequence [<SEQ ID 680>] (SEQ ID NO: 680):

25

30

35

40

45

50

		250	260	270	280	290	300
orf24-1.pep		SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
orf24ng		SILIPARVLPILMELHTISVVFIASGTERINTSSEGDIPFCTSAEKPPIKDTPMALAALS					
5		250	260	270	280	290	300
	orf24-1.pep	KVCATLT	X				
	orf24ng	KVCATLT	X				

- 10 Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

- 15 The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 681] (SEQ ID NO: 681):

20

```

1  ..ACCGACGTGC AAAAAGAGTT GGTCGGCGAA CAACGCAAGT GGGCGCAGGA
51  AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT
101 ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG
151 ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG

```

This corresponds to the amino acid sequence [SEQ ID 682; ORF25] (SEQ ID NO: 682; ORF25):

25

```

1  ..TDVQKELVGE QRKWAQEKIS NCRQAAAQAD RQEYAEYLKL QCDTRMTRE
51  IQYLRGYSID *

```

Further work revealed the complete nucleotide sequence [SEQ ID 683] (SEQ ID NO: 683):

30

```

1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
101 TGCAAGGCAT ACGCGGCAAT ATTCAGGAAA CGCTCAGCA GGAAGCGCGT
151 TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG ACAAATTAT
201 CGCCGCGGCC TACGTTTGG CGTTTCTTT GGAACACGCT TCGGAAACGC
251 AGGAAGGCGG GCGCACGTTT TGTATCGCCG ATTTGAACAT TACCGTGCCG
301 TCTGAAACGC TTGCCGATGC CAAGGCAAC AGCCCCTGT TGTACGGGGA
35  AACTGCTTTG TCGGATATTG TGCGGCAGAA GACGGGCGGC AATGTCGAGT
401 TTAAAGACGG CGTATTGACG GCAGCCGTCC GCTTCCTGCC CGTCAAAGAC
451 GGTCTAGACGG CATTGTGCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
501 GTCTGCCGCG CTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT GAAAAAAGAA GACGCGGTCA GGATTTTGAG CGGAAAAGCC
40  CGTGAAGAAG AACCGTCCAA ACCCACGCCC GAAGACATT TGGAAACAAA
601 TGCCGCCGGC GCGGATGCGG GCGTACCCCA AGCCGAGAA GGCGCGCCCG
651 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
701

```


5

751	GTATCACGGG	GCGAAGTGGG	AGAGGCGCGC	GTACAAAACC	AGCGTGCGGA
801	ATCCGAAATT	ACCAAACTTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
851	AGTTGGTCGG	CGAACAACGC	AAGTGGGCGC	AGGAAAAAAT	CAGCAACTGC
901	CGACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
951	GCTGCAATGC	GACACGCGGA	TGACGCGCGA	ACGGATACAG	TATCTTCGCG
1001	GCTATTCCAT	GATTAG			

This corresponds to the amino acid sequence [[<SEQ ID 684; ORF25-1>](#)] ([SEQ ID NO: 684; ORF25-1](#)):

10	1	MYRKLIALPF	ALLLAACGRE	EPPKALECAN	PAVLQGIRGN	IQETLTQEAR
	51	SFAREDGRQF	VDADKIIAAA	YGLAFSLEHA	SETQEGGRTF	CIADLNITVP
	101	SETLADAKAN	SPLLYGETAL	SDIVRQKTGG	NVEFKDGVLT	AAVRFLPVKD
	151	GQTAFDVNTV	GMAAQTLTAA	LLPYGVKSIV	MIDGKAVKKE	DAVRILSGKA
	201	REEEPSKPTP	EDILEHNAAG	GDAGVPQAAE	GAPEPEILHP	DDGERADTVT
15	251	VRSRGEVEEAR	VQNQRAESEI	TKLWGGGLTD	VQKELVGEQR	KWAQEKISNC
	301	RQAAQADRO	EYAEYLKLOC	DTRMTREIRQ	YLRGYSID*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF25 (SEQ ID NO: 682) shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) (SEQ ID NO: 686) from strain A of *N. meningitidis*:

```

                                10      20      30
orf25.pep                      TDVQKELVGEQQRKWAQEKISNCRQAAAQAD
                                |||
25 orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEXRKWAQEKISNCRQAAAQAD
                   250      260      270      280      290      300

                                40      50      60
orf25.pep      RQEYAEYKLQCDTRMTRERIQYLRGYSIDX
                                |||
30 orf25a      RQEYAEYKLQCDTRMTRERIQYLRGYSIDX
                   310      320      330

```

The complete length ORF25a nucleotide sequence [<SEQ ID 685>] (SEQ ID NO: 685) is:

35	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAANGCAT	ACGCNGCAAT	ATTCAGGAAA	CGCTACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACNG	CANGCAGTTT	GTGATGCCG	ACNAAATTAT
	201	CGCCGCCGCC	TANGNTNNGN	NGNTNTCTTT	GGAACACGCT	TCGGAAACGC
40	251	AGGAAGGCGG	GCGCACGTTT	TGTNTCGCCG	ATTTGAACAT	TACCGTGGCG
	301	TCTGAACCGC	TTGCCGATGC	CAAGGCAAAC	AGCCCCCTGC	TAGCGGGGA
	351	AACCGCTTTG	TCGGATATTG	TGCGGCAGAA	GACGGGCGGC	AATGTCTGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCTTACC	CGTCAAAGAC
	451	GGTCAGANGG	CATTTGTGCA	CAACACGTC	GGTATGGCGG	CGCAAACGCT
45	501	GTCTGCCCGC	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	AAAAAAGAA	GACGCGGTCA	GGATTNTGAG	CNGANAAGCC
	601	CGTGAANAAG	AACCGTCCAA	ANCCNNGCCC	GAGAACATTT	TGGAACATGAA
	651	TGCCGCCCGA	GGGGATGCAG	ACGTACCCCA	AGCCGGAGAA	GACGCGCCCC

5
 701 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
 751 GTATCACGGG GCGAAGTGGA AGAGGCGCGN GTACAAAACC AGCGTGCGGA
 801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
 851 AGTTGGTCGG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
 901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
 951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
 1001 GCTATTCCAT CGATTAG

This encodes a protein having amino acid sequence [SEQ ID 686] (SEQ ID NO: 686):

10
 1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
 51 SFAREDXXQF VDADXI IAAA XXXXSLEHA SETQEGGRTF CXADLNITVP
 101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
 151 GQXAFVDNTV GMAAQTL SAA LLPYGVKSIV MIDGKAVKKE DAVRIXSXXA
 201 REXEPSKXXP EDILEHNAAG GDADVPOAGE DAPEPEILHP DDGERADTVT
 15 251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEXR KWAQEKISNC
 301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

ORF25a (SEQ ID NO: 686) and ORF25-1 (SEQ ID NO: 684) show 93.5% identity in 338 aa overlap:

20
 orf25a.pep 10 20 30 40 50 60
 MYRKLIALPFALLLAACGRE EPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF
 orf25-1 MYRKLIALPFALLLAACGRE EPPKALECANPAVLQGIRGNIQETLTQEARSFAREDGRQF
 10 20 30 40 50 60

25
 orf25a.pep 70 80 90 100 110 120
 VDADXI IAAAXXXXSLEHASETQEGGRTFCXADLNITVPSETLADAKANSPLLYGETAL
 orf25-1 VDADKI IAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAKANSPLLYGETAL
 70 80 90 100 110 120

30
 orf25a.pep 130 140 150 160 170 180
 SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQXAFVDNTVGMAAQTL SAA LLPYGVKSIV
 orf25-1 SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQTAFVDNTVGMAAQTL SAA LLPYGVKSIV
 130 140 150 160 170 180

35
 orf25a.pep 190 200 210 220 230 240
 MIDGKAVKKEDAVRIXSXXAREXEPSKXXPEDILEHNAAGGDADVPOAGEDAPEPEILHP
 orf25-1 MIDGKAVKKEDAVRILSGKAREE E P S K P T P E D I L E H N A A G G D A G V P Q A A E G A P E P E I L H P
 190 200 210 220 230 240

40
 orf25a.pep 250 260 270 280 290 300
 DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNC
 orf25-1 DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNC
 250 260 270 280 290 300

45
 orf25a.pep 310 320 330 339
 RQAAAQADRQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
 orf25-1 RQAAAQADRQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
 310 320 330

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 (SEQ ID NO: 682) shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) (SEQ ID NO: 688) from *N.gonorrhoeae*:

```

5      orf25.pep                                TDVQKELVGEQRKWAQEKISNCRQAAAQAD    30
      orf25ng      VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNCRQAAAQAD    308

      orf25.pep      RQEYAEYLKLQCDTRMTRERIQYLRGYSID    60
      orf25ng      RQEYAEYLKLQCDTRMTRERIQYLRGYSID    338

```

The complete length ORF25ng nucleotide sequence [<SEQ ID 687>] (SEQ ID NO: 687) is:

```

1  ATGTATCGGA  AACTCATTCG  GCTGCCGTTT  GCCCTGCTGC  TTGCAGCGTG
51 CGGCAGGGAA  GAACCGCCCA  AGGCGTTGGA  ATGCGCCAAC  CCCGCCGTGT
101 TGCAGGACAT  ACGCGGCAGT  ATTCAGGAAA  CGCTCACGCA  GGAAGCGCGT
151 TCTTTCGCGC  GCGAAGACGG  CAGGCAGTTT  GTCGATGCCG  ACAAATTAT
201 CGCCGCCGCC  TACGTTTGG  CGTTTCTTT  GGAACACGCT  TCGAAACGC
251 AGGAAGCGCG  GCGCACGTTT  TGTATCGCCG  ATTTGAACAT  TACCGTGCCG
301 TCTGAAACGC  TTGCCGATGC  CGAGGCAAA  AGCCCCCTGC  TGTATGGGGA
351 AACGTCTTTG  GCAGACATCG  TGCAGCAGAA  GACGGGCGGC  AATGTCGAGT
20  401 TTAAAGACGG  CGTATTGACG  GCAGCCGTCC  GCTTCCTGCC  CGCCAAAGAC
451 GCTCGGACGG  CATTTATCGA  CAACACGGTC  GGTATGGCGA  CGCAAACGCT
501 GTCTGCCGCG  TTGCTGCCTT  ACGGCGTGAA  GAGCATCGTG  ATGATAGACG
551 GCAAGGCGGT  GACAAAAGAA  GACGCGGTCA  GGGTTTGTAG  CGGCAAAGCC
601 CGTGAAGAAG  AACCGTCCAA  ACCCACCCCC  GAAGACATTT  TGGAACACAA
25  651 TGCCGCCGCG  GGCATGCGCG  GCGTACCCCA  AGCCGCAGAA  GCGCACCCCG
701 AACCCGAAAT  CCTGCATCCC  GACGACGTCG  AGCGTGCCGA  TACCGTTACC
751 GTATCACGGG  GCGAAGTGGA  AGAGGCGCGC  GTACAAAACC  AACGTGCGGA
801 ATCCGAAATT  ACCAACTTT  GGGGAGGACT  CGATACCGAC  GTGCAAAAAG
851 AGTTGGTCGG  CGAACAGCGC  AAGTGGGCGC  AGGAAAAAAT  CAGcaactgc
30  901 cgACAAGCCG  CCGCGCAGGC  AGACCGCAG  GAATACGCCG  AATACCTCAA
951 GCTCCAATGC  GACACGCGGA  TGACGCGCGA  AcggaTACAG  TATCTTCGCG
1001 GCTATTCCAT  CGATTAG

```

This encodes a protein having amino acid sequence [<SEQ ID 688>] (SEQ ID NO: 688):

```

35  1  MYRKLIALPF  ALLLAACGRE  EPPKALECAN  PAVLQDIRGS  IQETLTQEAR
51  SFAREDGRQF  VDADKIIAAA  YGLAFSLEHA  SETQEGGRTP  CIADLNITVP
101 SETLADAEAN  SPLLYGETSL  ADIVQKTGG  NVEFKDGLT  AAVRFLPAKD
151 ARTAFIDNTV  GMATQTLASA  LLPYGVKSIV  MIDGKAVTKE  DAVRVLSGKA
201 REEPPSKPTP  EDILEHNAAG  GDAGVPQAAE  GAPEPEILHP  DDVERADTVT
40  251 VSRGEVEEAR  VQNQRAESEI  TKLWGGLDTD  VQKELVGEQR  KWAQEKISNC
301 RQAAAQADRQ  EYAEYLKLQC  DTRMTRERIQ  YLRGYSID*

```

ORF25ng (SEQ ID NO: 688) and ORF25-1 (SEQ ID NO: 684) show 95.9% identity in 338 aa overlap:

```

45      orf25-1.pep      MYRKLIALPFALLLAACGRE EPPKALECANPAVLQGIRGNIQETLTQEARSFAREDGRQF
      |||:|||||

```

		orf25ng	MYRKLIALPFALLLAACGREEPPKALECANPAVLQDIRGSIQETLTQEARSFAREDGRQF	10	20	30	40	50	60
				70	80	90	100	110	120
5		orf25-1.pep	VDADKIIAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAKANSPLLYGETAL						
		orf25ng	VDADKIIAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAEANSPLLYGETSL						
				70	80	90	100	110	120
10		orf25-1.pep	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGTAFVDNTVGMAAQTLAALLPYGVKSIV	130	140	150	160	170	180
		orf25ng	ADIVQQKTGGNVEFKDGVLTAAVRFLPAKDARTAFIDNTVGMATQTLAALLPYGVKSIV						
				130	140	150	160	170	180
15		orf25-1.pep	MIDGKAVKKEDAVRILSGKAREEESP KPTPEDILEHNAAGGDAGVPQAAGAPEPEILHP	190	200	210	220	230	240
		orf25ng	MIDGKAVTKEDAVRVLSGKAREEESP KPTPEDILEHNAAGGDAGVPQAAGAPEPEILHP						
				190	200	210	220	230	240
20		orf25-1.pep	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNC	250	260	270	280	290	300
		orf25ng	DDVERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNC						
				250	260	270	280	290	300
25		orf25-1.pep	RQAAAQADRQEYAEYLKLQCDTRMTRERIQYLRGYSIDX	310	320	330	339		
		orf25ng	RQAAAQADRQEYAEYLKLQCDTRMTRERIQYLRGYSIDX						
				310	320	330			

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (SEQ ID NO: 684) (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 (SEQ ID NO: 684) is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1 (SEQ ID NO: 684).

Example 82

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 689] (SEQ ID NO: 689)

```

1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
101 GCATCGGTAT TCTGGwysGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
201 CGsyGATTGG TCGCTGGGCA AACCAGAAAT CTTGGTTTTT CkGATACTTT
251 TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....

//

851 .....AC TTCGCTGGTA
901 TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
151 GATTAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
1001 TGTTTCGGCGC AATCGCAATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
1051 GTCGGCGAAA TGCACACGGG CGATTACCTC TCCACACTGG TTGCGGGCAA
1101 CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGTGCCG
201 ATTGCCCCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCGCCGCG
1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
251 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
1501 AAAAAA..

```

This corresponds to the amino acid sequence [SEQ ID 690; ORF26] (SEQ ID NO: 690; ORF26):

```

30 1  MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFVGGNPV
51  DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSL LTYSGSN...

//

251 .....TSLV
301 FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
35 351 VGEMHTGDYL STLVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARNHI
451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
501 KK..

```

Further work revealed the complete nucleotide sequence [SEQ ID 691] (SEQ ID NO: 691):

```

1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
101 GCATCGGTAT TCTGGTCGGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
45 201 CCGCGATTGG TCGCTGGGCA AACCAGAAAT CTTGGTTTTT CTGATACTTT
251 TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT

```

5
10
15
20
25

```

301  GCCGACTGGG  CAAAACGGCA  CATTAAAAAC  CGGCGCGGCG  CGAAAATGCT
351  GACCGCCTGC  CTCGTGTTTC  TAACCTTTAT  CGACGACTAT  TTCCACAGTC
401  TCGCCGTCGG  TCGGATTGCC  CGCCCCGTTA  CCGACAAGTT  TAAAGTTTCC
451  CGCACCAAAC  TCGCTACAT  CCTCGACTCC  ACTGCCGCTC  CTATGTGCGT
501  GCTGATGCCC  GTTCAAGCT  GGGGCGCGTC  GATTATCGCC  ACGCTTGCCG
551  GACTGCTCGT  TACCTACAAA  ATCACCGAAT  ACACGCCGAT  GGGGACGTTT
601  GTCGCCATGA  GCCTGATGAA  CTATTACGCA  CTGTTTGCCC  TGATTATGGT
651  GTTCGTGCTC  GCATGGTTTT  CCTTCGACAT  CGGCTCGATG  GCACGTTTCG
701  AACAAAGCCG  GTTGAACGAA  GCCCAGCATG  AAACGCGCT  TTCAGACGCT
751  ACCAAAGGTC  GTGTTTACGC  ACTGATTATT  CCCGTTTTGG  CCTTAATCGC
801  CTCAACGGTT  TCCGCCATGA  TCTACACCGG  CGCGCAGGCA  AGCGAAACCT
851  TCAGCATTTT  GGGGGCATTT  GAAAACACGG  ACGTAAACAC  TTCGCTGGTA
901  TTCGGCGGCA  CTTGCGGCGT  CCTTGCCGTC  GTTCTCTGCA  CGCTCGGCAC
951  GATTAAAAAC  GCCGACTATC  CCAAAGCCGT  TTGGCAGGGT  GCGAAATCTA
1001  TGTTCGGCGC  AATCGCCATT  TTAATCCTCG  CTTGGCTCAT  CAGTACGGTT
1051  GTCGGCGAAA  TGCACACCGG  CGATTACCTC  TCCACACTGG  TTGCGGGCAA
1101  CATCCATCCC  GGCTTCCTGC  CCGTCATCCT  CTTCTGCTC  GCCAGCGTGA
1151  TGGCGTTTGC  CACAGGCACA  AGCTGGGGGA  CGTTCGGCAT  TATGCTGCCG
1201  ATTGCCGCCG  CCATGGCGGT  CAAAGTCGAA  CCCGCGCTGA  TTATCCCGTG
1251  TATGTCCGCA  GTAATGGCGG  GGGCGGTATG  CGGCGACCAC  TGCTCGCCCA
1301  TTTCCGACAC  GACCATCCTG  TCGTCCACCG  GCGCGCGCTG  CAACCACATC
1351  GACCACGTTA  CCTCGCAACT  GCCTTACGCC  TTAACCGTTG  CCGCCGCCGC
1401  CGCATCGGGC  TACCTCGCAT  TGGGTCTGAC  AAAATCCGCG  CTGTTGGGCT
1451  TTGGCACGAC  AGGCATTGTA  TTGGCGGTGC  TGATTTTTCT  GTTGAAAGAT
1501  AAAAAACGCG  CCAACGCCTG  A

```

This corresponds to the amino acid sequence [<SEQ ID 692; ORF26-1>] (SEQ ID NO: 692; ORF26-1):

30
35
40

```

1  MQLIDYSHSF  FSVVPPFLAL  ALAVITRRVL  LSLGIGILVG  VAFLVGGNPV
51  DGLTHLKDMV  VGLAWSGDGW  SLGKPKILVF  LILLGIFTSL  LTYSGSNQAF
101  ADWAKRHIKN  RRGAKMLTAC  LVEVTFIDDY  FHSLAVGAIA  RPYTDKFKVS
151  RTKLAYILDS  TAAPMCVLMF  VSSWGASIIA  TLAGLLVITYK  ITEYTPMGTF
201  VAMSLMNYIA  LFAIMVFVV  AWFSFDIGSM  ARFEQAALNE  AHDETAVSDA
251  TKGRVYALII  PVLALIASTV  SAMIYTGAQA  SETFSILGAF  ENTDVNTSLV
301  FGGTCGVLA  VLCTLGTIKT  ADYPKAVWQG  AKSMFGAIAI  LILAWLISTV
351  VGEMHTGDYL  STLVAGNIHP  GFLPVILFLL  ASVMAFATGT  SWGTFGIMLP
401  IAAAMAVKVE  PALIIPCMSA  VMAGAVCGDH  CSPISDTTIL  SSTGARCNIH
451  DHVTSQLPYA  LTVAAAAASG  YLALGLTKSA  LLGFGTGTGIV  LAVLIFLLKD
501  KKRANA*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 (SEQ ID NO: 1156) of *H.influenzae* (accession number P44263)

45 ORF26 (SEQ ID NO: 690) and HI1586 (SEQ ID NO: 1156) show 53% and 49% amino acid identity in 97 and 221 aa overlap at the N-terminus and C-terminus, respectively:

```

Orf26 1  MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60
      M+LID+S S +S+VP LA+ LA+ TRRV L +L V
HI1586 14 MELIDFSSSVWSIVPALLAIILAIATRRVLVLSAGIIIGSLMLSVDWQIGSAFNLYLVKNV 73

```

-511-

Orf26 61 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTTYSGSN 97
 V L ++D + + I++F +LLG+ T+LLT SGSN
 HI1586 74 VSLVYADGEIN-SNMNIVLFLLLLGLVLTALLTVSGSN 109

//

5 Orf26 86 IFTSLTTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGA KSMFGXXXX 141
 +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
 HI1586 299 VFSVLGTFTENTVVGTSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGSAIAI 358

Orf26 142 XXXXXXSTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGTSGWTFGIMLP 201
 + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP
 10 HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSGWTFGIMLP 418

Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISTTILSSTGARNHIDHVT SQXXXX 261
 IAAAMA P L++PC+SAVMAGAVCGDHCSPISTTILSSTGA+CNHIDHVT+Q
 HI1586 419 IAAAMAANAPELLLPCLSAVMAGAVCGDHCSPISTTILSSTGAKCNHIDHVTQLPYA 478

Orf26 262 XXXXXXXXXXXXXXXXXXXKSALLGFGTTGIVLAVLIFLLKDK 302
 S L GF T + L V+IF +K +
 15 HI1586 479 ATVATATSIGYIVVGFTYSGLAGFAATAVSLIVIIFAVKKR 519

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF26 (SEQ ID NO: 690) shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) (SEQ ID NO: 694) from strain A of *N. meningitidis*:

20 orf26.pep 10 20 30 40 50 60
 MQLIDYSHSFFSVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV
 |||||
 orf26a MQLIDYSHSFFSVPPFLALALAVITRRVLLSLGIGILVGVAFVGGNPVDGLTHLKDMV
 10 20 30 40 50 60

25 orf26.pep 70 80 90 99
 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTTYSGSNXX-----
 |||||
 orf26a VGLAWSGDGWSLGKPKXLVFLILLGIFTSLTTYSGSNQAFADWAKRHIKNRRGAKMLTAC
 70 80 90 100 110 120

30 orf26.pep -----
 orf26a LVFVTFIDYFHS LAVGAXARPVTDKFKVSRKLAYILDSTAAPMCVLMPVSSWGASIIA
 130 140 150 160 170 180

35 orf26.pep -----
 orf26a TLAGLLV TYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSDIGSMARFEQAALNE
 190 200 210 220 230 240

40 orf26.pep ----- 100 110
 -----TSLV
 |||||
 orf26a AHDETAVSDGSGR VYALIIPVLALIASTVSAMIYTGAAQASETFSILGAFENTDVNTSLV
 250 260 270 280 290 300

		120	130	140	150	160	170
	orf26.pep	FGGTCGVFAVLCTLGTIKTADYPKAVWQGA SMFGAIAILILAWLISTVV GEMHTGDYL					
	orf26a	FGGTCGVLA VVLCTLG TIKIADYPKAVWQGA SMFGAIAILILAWLISTVV GEMHTGDYL					
5		310	320	330	340	350	360
		180	190	200	210	220	230
	orf26.pep	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPC MSA					
	orf26a	STLVAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVDPSLIIPC MSA					
10		370	380	390	400	410	420
		240	250	260	270	280	290
	orf26.pep	VMAGAVCGDH CSPISD TTILSSTGARNHIDH VTSQLPYALT VAAAAASGYLALGLTKSA					
	orf26a	VMAGAVCGDH CSPISD TTILSSTGARNHIDH VTSQLPYALT VAAAAASGYLALGLTKSA					
15		430	440	450	460	470	480
		300	310				
	orf26.pep	LLGFGTTGIVLAVLIFLLKDKK					
	orf26a	LLGFGXTGIVLAVLIFLLKDKKRANAX					
20		490	500				

The complete length ORF26a nucleotide sequence [<SEQ ID 693>] (SEQ ID NO: 693) is:

	1	ATGCAGCTGA	TCGACTATTC	ACATTCATTT	TTCTCGGTTG	TGCCACCCTT
	51	TTTGGCACTG	GCACTTGCCG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
25	101	GCATCGGTAT	TCTGGTCGGC	GTTGCCTTTT	TGGTCGGCGG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGTCAGA
	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAANT	CTTGGTTTTC	CTGATACTTT
	251	TGGGTATTTT	TACTTCCCTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
	301	GCCGACTGGG	CAAAACGGCA	CATTAAAAAC	CGGCGCGGCG	CGAAAATGCT
30	351	GACCGCCTGC	CTCGTGTTCG	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
	401	TCGCCGTCGG	TGCGNTTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TCGCCTACAT	CCTCGACTCC	ACTGCCGCGC	CTATGTGCGT
	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
	551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACACGCCGAT	GGGGACGTTT
35	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGGT
	651	GTTTCGTCGTC	GCATGGTTCT	CCTTCGACAT	CGGCTCGATG	GCACGTTTCG
	701	AACAAGCCGC	GTTGAACGAA	GCCCACGATG	AAACTGCCGT	TTCAGACGGC
	751	AGCTGGGGCA	GGGTTTACGC	ATTGATTATT	CCCGTTTGGG	CCTTAATCGC
	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	TGCACAGGCA	AGCGAAACCT
40	851	TCAGCATTTT	GGGTGCATTT	GAAAATACGG	ACGTGAACAC	TTGCTGGGTA
	901	TTGCGCGGCA	CTTGCGGCGT	GCTTGCCGTC	GTCCTCTGCA	CGCTCGGCAC
	951	GATTAAAAATC	GCCGATTATC	CCAAAGCCGT	TTGGCAGGGT	GCGAAATCCA
	1001	TGTTTCGGCGC	AATCGCCATT	TTAATCCTTG	CCTGGCTCAT	CAGTACGGTT
	1051	GTCGGCGAAA	TGCACACAGG	CGACTACCTC	TCCACGCTGG	TTGCGGGCAA
45	1101	CATCCATCCC	GGCTTCTTGN	CCGTCATCCT	TTTCTTGCTC	GCCAGCGTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCGGCAT	CATGTGCCCG
	1201	ATTGCCGCGG	CCATGGCGGT	CAAAGTCGAT	CCCTCACTGA	TTATCCCGTG
	1251	TATGTCCGCC	GTGATGGCGG	GGGCGGTATG	CGGCGACCAC	TGCTCGCCCA
	1301	TTTCCGACAC	GACCATCCTG	TCGTCCACCG	GCGCGCGCTG	CAACCACATC
50	1351	GACCAAGTTA	GNTCGCAACT	GCCTTACGCC	TTAACCCTTG	CCGCCGCCGC
	1401	CGCATCGGGN	TACCTCGCAT	TGGGTCTGAC	AAAATCCGCG	CTGTTGGGTT
	1451	TTGGCANGAC	AGGCATTGTA	TTGGCGGTGC	TGATTTTCT	GTTGAAAGAT
	1501	AAAAAACGCG	CCAACGCTG	A		

55 This encodes a protein having amino acid sequence [<SEQ ID 694>] (SEQ ID NO: 694):

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV
 51 DGLTHLKDMV VGLAWSDDGW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAXA RPVTDKFKVS
 151 RAKLAYILDS TAAPMCVLMP VSSWGASIIA TLAGLLVITYK ITEYTPMGTF
 201 VAMSLMNYA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHDETAUSDG
 251 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLAV VLCTLGTIKI ADYPAVWQG AKSMFGAIAI LILAWLISTV
 351 VGEMHTGDYL STLVAAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVD PSLIIPCMA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
 451 DHVTSQLPYA LTVAASAAAG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD
 501 KKRANA*

ORF26a (SEQ ID NO: 694) and ORF26-1 (SEQ ID NO: 692) show 97.8% identity in 506 aa overlap:

15		10	20	30	40	50	60
	orf26a.pep	MQLIDYSHSFSSVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV					
	orf26-1	MQLIDYSHSFSSVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV					
		10	20	30	40	50	60
20		70	80	90	100	110	120
	orf26a.pep	VGLAWSDDGWSLGKPKXLVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC					
	orf26-1	VGLAWSDDGWSLGKPKILVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC					
		70	80	90	100	110	120
25		130	140	150	160	170	180
	orf26a.pep	LVFVTFIDYFHSLAVGAXARPVTDKFKVSRALAYILDSTAAPMCVLMPVSSWGASIIA					
	orf26-1	LVFVTFIDYFHSLAVGAIARPVTDKFKVSRALAYILDSTAAPMCVLMPVSSWGASIIA					
		130	140	150	160	170	180
30		190	200	210	220	230	240
	orf26a.pep	TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE					
	orf26-1	TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE					
		190	200	210	220	230	240
35		250	260	270	280	290	300
	orf26a.pep	AHDETAUSDGWSGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV					
	orf26-1	AHDETAUSDATKGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV					
		250	260	270	280	290	300
40		310	320	330	340	350	360
	orf26a.pep	FGGTCGVLAVVLCTLGTIKIADYPAVWQGA KSMFGAIAI LILAWLISTVVGEMHTGDYL					
	orf26-1	FGGTCGVLAVVLCTLGTIKTADYPAVWQGA KSMFGAIAI LILAWLISTVVGEMHTGDYL					
		310	320	330	340	350	360
45		370	380	390	400	410	420
	orf26a.pep	STLVAGNIHPGFLXVILFLLASVMAFATGT SWGTFGIMLP IAAAMAVKVDPSLIIPCMA					
	orf26-1	STLVAGNIHPGFLPVILFLLASVMAFATGT SWGTFGIMLP IAAAMAVKVEPALIIPCMA					
		370	380	390	400	410	420

-514-

		430	440	450	460	470	480
	orf26a.pep	VMAGAVCGDHCSPISD	TTILSSTGARC	NHIDHVTSQ	LPYALTVA	AAAAASGY	LALGLTKSA
5	orf26-1	VMAGAVCGDHCSPISD	TTILSSTGARC	NHIDHVTSQ	LPYALTVA	AAAAASGY	LALGLTKSA
		430	440	450	460	470	480
		490	500				
	orf26a.pep	LLGFGXTGIVLAVLI	FLLKDKKRANAX				
10	orf26-1	LLGFGTTGIVLAVLI	FLLKDKKRANAX				
		490	500				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 (SEQ ID NO: 690) shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) (SEQ ID NO: 696) from *N. gonorrhoeae*:

15	orf26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLSLIGIGILXXVAFLVGGNPVDGLTHLKDMV	60
	orf26ng	MQLIDYSHSFFSVVPPFLALALAVITRRVLSLIGIGILVGVAFVGGNPVDGLTHLKDMV	60
	orf26.pep	VGLAWSDXDWSLGGPKILVFXILLGIFTSLLTYSGSN	97
20	orf26ng	VGLAWADGDWSLGGPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
		//	
	orf26.pep	TSLVFGGTCGVFAVVLCTLGTIKTADYPKA	326
	orf26ng	ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVLA VVLCTFGTIKTADYPKA	326
25	orf26.pep	VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF	386
	orf26ng	VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF	386
	orf26.pep	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISD TTILSSTGAR	446
30	orf26ng	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISD TTILSSTGAR	446
	orf26.pep	CNHIDHVTSQ LPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLI FLLKDKK	502
	orf26ng	CNHIDHVTSQ LPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLI FLLKDKKRADV	506

35 The complete length ORF26ng nucleotide sequence [<SEQ ID 695>] (SEQ ID NO: 695) is:

	1	ATGCAGCTGA	TTGACTATTC	ACATTCATTT	TTCTCGGTTG	TGCCACCCTT
	51	TTTG GCACTG	GCACTTGCCG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
	101	GCATCGGTAT	TTTG GTCGGC	GTTGCCTTTT	TGGTCGGCGG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGGCAGA
40	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAAAT	CTTGGTTTTT	CTGATACTTT
	251	TGGGCATTTT	CAC TTCACTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
	301	GCCGACTGGG	CAAAACGGCA	CATTAAAAAC	CGGTGCGGCG	CGAAAATGCT
	351	GACCGCTGCG	CTCGTGTTCTG	TAACCTTTAT	CGACGACTAT	TTCCACAGCC

-515-

5
10
15
20
25

```

401 TCGCCGTCGG TGCATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
451 CGCGCCAAAC TCGCCTACAT CCTCGACTCC ACTGCCTCGC CCATGTGCGT
501 GCTGATGCCC GTTTCAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
551 GATTGCTCGT TACCTACAAA ATTACCGAAT ACACGCCGAT GGGGACGTTT
601 GTCGCCATGA GCCTGATGAA CTATTACGCG CTGTTTGCCC TGATTATGGT
651 ATTCGTCGTC GCATGGTTCT CCTTCGACAT CCGCTCGAtg gCGGTTTTCG
701 AACAGGCTGC GTTGAACGAA gcccaggacg aaaccgcccgc tTCAGACgCT
751 ACCAAAGGTC GTGTTTACGC ATTGATTATT CCCGTTTGG CCTTAATCGC
801 CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
851 TCAGCATTTT GGGGGCATTT GAAAATACCG ACGTAAACAC TTCGCTGGTA
901 TTCGGCGGCA CTTCGGCGCT GCTTGCCGTC GTCCTCTGCA CGTTCGGCAC
951 GATTAAAACC GCCGATTATC CCAAAGCCGT GTGGCAGGGT GCGAAATCCA
1001 TGTTCCGCGC AATCGCCATT TTAATCCTCG CCTGGCTCAT CAGTACGGTT
1051 GTCGGCGAAA TGCACACGGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
1101 CATCCATCCC GGCTTCTCTG CCGTCATCCT CTTCCTGCTC GCCAGCGTGA
1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
1201 ATTGCCGCGG CCATGCGCGT CAAAGTCGAA CCCGCGCTGA TTAtcccGTG
1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGTTCGCCCA
1301 TCTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
1351 GACCACGTTA CCTCGCAACT GCCTTATGCC CTGACGGTTG CCGCCGCCGC
1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
1451 TTGGCACGAC CGGTATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
1501 AAAAAACGCG CCGACGTTTG A

```

25 This encodes a protein having amino acid sequence [SEQ ID 696] (SEQ ID NO: 696):

30
35

```

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV
51 DGLTHLKDMV VGLAWADGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
101 ADWAKRHIKN RCGAKMLTAC LVFVTFIDDY FHSLAVGAIA RPVTDKFKVS
151 RAKLAYILDS TASPMCVLMP VSSWGASIIA TLAGLLVITYK ITEYTPMGTF
201 VAMSLMNYIA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AQDETAASDA
251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
301 FGGTCGVLAV VLCTFGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
351 VGEHMTGDYL STLVAGNIHP GFLPVILFLL ASVMFATGT SWGTFGIMLP
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
501 KKRADV*

```

ORF26ng (SEQ ID NO: 696) and ORF26-1 (SEQ ID NO: 692) show 98.4% identity in 505 aa overlap:

40
45
50

```

      10      20      30      40      50      60
orf26-1.pep MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV DGLTHLKDMV
          |||
orf26ng     MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV DGLTHLKDMV
          |||
      10      20      30      40      50      60

      70      80      90     100     110     120
orf26-1.pep VGLAWSGDW SLGKPKILV FLILLGIFTSL LTYSGSNQAF ADWAKRHIKN RRGAKMLTAC
          |||
orf26ng     VGLAWADGW SLGKPKILV FLILLGIFTSL LTYSGSNQAF ADWAKRHIKN RCGAKMLTAC
          |||
      70      80      90     100     110     120

      130     140     150     160     170     180
orf26-1.pep LVFVTFIDDY FHSLAVGAIA RPVTDKFKV SRKLA YILDSTAAPMCV LMPVSSWGASIIA
          |||
orf26ng     LVFVTFIDDY FHSLAVGAIA RPVTDKFKV SRAKLA YILDSTAAPMCV LMPVSSWGASIIA
          |||

```

-516-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf26-1.pep	TLAGLLVTYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE					
	orf26ng	TLAGLLVTYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE					
		190	200	210	220	230	240
		250	260	270	280	290	300
10	orf26-1.pep	AHDETAVSDATKGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV					
	orf26ng	AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV					
		250	260	270	280	290	300
		310	320	330	340	350	360
15	orf26-1.pep	FGGTCGVLAVVLCITLGTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTGDYL					
	orf26ng	FGGTCGVLAVVLCITLGTIKTADYPKAVWQGAQSMFGAIAILILAWLISTVVGEMHTGDYL					
		310	320	330	340	350	360
		370	380	390	400	410	420
20	orf26-1.pep	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALIIPCMSA					
	orf26ng	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALIIPCMSA					
		370	380	390	400	410	420
		430	440	450	460	470	480
25	orf26-1.pep	VMAGAVCGDHCSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA					
	orf26ng	VMAGAVCGDHCSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA					
		430	440	450	460	470	480
		490	500				
30	orf26-1.pep	LLGFGTTGIVLAVLIFLLKDKKRANAX					
	orf26ng	LLGFGTTGIVLAVLIFLLKDKKRADVX					
		490	500				

In addition, ORF26 ng (SEQ ID NO: 696) shows significant homology to a hypothetical *H. influenzae* protein (SEQ ID NO: 1156):

35	sp P44263 YF86_HAEIN HYPOTHETICAL PROTEIN HI1586)gi 1074850 pir C64037
	hypothetical
	protein HI1586 - Haemophilus influenzae (strain Rd KW20))gi 1574427 (U32832) H.
	influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
40	Score = 538 bits (1370), Expect = e-152
	Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)
	Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXXAFVGGNPVDGLTHLKDMV 60
	M+LID+S S +S+VP LA+ LA+ TRR L +L V
	Sbjct: 14 MELIDFSSSVWSIVPALLAIILAIATRRVLVLSAGIIGSLMLS DWQIGSAFNLYVKNV 73
45	Query: 61 VGLAWADGDWSL GKPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC 120
	V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A
	Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLLGVLTALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132
	Query: 121 LVFVTFIDDYFHSLAVGAIARPVTDKFKVSRAKLAYILDSTASPMCVLMPVSSWGASIIA 180

L VFVTFIDDFHSLAVGAIARPVTD+FKVSRKLAYILDSTA+PMCV+MPVSSWGA II
 Sbjct: 133 LVFVTFIDDFHSLAVGAIARPVTD+FKVSRKLAYILDSTAAPMCVMPVSSWGAYIIT 192
 Query: 181 TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFFVAVFSDIGSMARFEQAALNE 240
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL
 5 Sbjct: 193 LIGGLLATYSITEYTPIGAFVAMSSMNFYAIFSIIMVFFVAYFSDIASMVRHEKLALKN 252
 Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA----SETFSILGAFENTDVN 296
 +D+ TKG+V LI+P+L LI +TVS MIYTGA+A + FS+LG FENT V
 Sbjct: 253 TEDQLEETGTGKQVRNLILPILVLIATVSMMIYTGAEEALADGKVFSVLGTFENTVVG 312
 Query: 297 TSLVFGGTCGVL--AVVLCFTGTIKTADYPKAVWQGA KSMFGXXXXXXXXXXSTVVGEM 354
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M
 10 Sbjct: 313 TSLVGGGFCSIIISTLLIILDRQVSVEYVRSWIVGIKSMGAIAILFFAWTINKIVGDM 372
 Query: 355 HTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSWGTGFGIMLPAAAMAVKVEPALI 414
 TG YLS+LV+GNI FLPVILF+L + MAF+TGTSGWGTGFGIMLPAAAMA P L+
 Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTGFGIMLPAAAMAANAPELL 432
 Query: 415 IPCMSAVMAGAVCGDHCSPISTTILSSTGARNHIDHVT SQXXXXXXXXXXXXXXXXXXXX 474
 +PC+SAVMAGAVCGDHCSPISTTILSSTGA+CNHIDHVT+Q
 15 Sbjct: 433 LPCLSAVMAGAVCGDHCSPISTTILSSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492
 Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501
 S L GF T + L V+IF +K +
 20 Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 697>] (SEQ ID NO: 697):

1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
 51 CGATGAGCCT GCCAAATTC TGA CTGGGA TGAAAGCGGC CGATTACTCT
 101 CGGA ACTGTC TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
 30 201 GGTCAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCCTGA

This corresponds to the amino acid sequence [<SEQ ID 698; ORF27>] (SEQ ID NO: 698; ORF27):

1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW
 35 51 YEDGSKKSEX VYQDDKLVRK TQWDKGYLI EP*

Further work revealed the complete nucleotide sequence [<SEQ ID 699>] (SEQ ID NO: 699):

1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC

5
10
15

```

51  GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA
101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT
401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA
501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT
601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
651 TTCTAAAAAG AGCGAAGCTG TTATCAGGA TGACAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCTGA

```

This corresponds to the amino acid sequence [<SEQ ID 700; ORF27-1>] (SEQ ID NO: 700; ORF27-1):

20

```

1  MKKLSRIVFS TVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIRQYSV
51  VAGIAHAQDF YPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
101 KMAAGGFSK GK PDGEWVNWYP NGKKSAMVMPY KNGLSEGTGY RYYRNGGKES
151 EIQFKQNKAN GVKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS
201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

```

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF27 (SEQ ID NO: 698) shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) (SEQ ID NO: 702) from strain A of *N. meningitidis*:

30

```

      orf27.pep                                10      20      30
      KQWYADXSIKTEMVMVNDEPAKILTWDESG
      ||||| : |||||
      orf27a      LSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVNDEPAKILTWDESG
                  140      150      160      170      180      190

      40      50      60      70      80
      orf27.pep      RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDGYLIEPX
      ||||| : |||||
      orf27a      RLLSELSIRHHQRNGVVLEWYEDGSKKXEA VYQDDKLVRKTQWDKDGYLIEPX
                  200      210      220      230      240

```

The complete length ORF27a nucleotide sequence [<SEQ ID 701>] (SEQ ID NO: 701) is:

40
45

```

1  ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
51  GGCCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAT CAGAACGGGA
101 AACTGACGGC GACGNTGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA
201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA
301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT

```

5

```

401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAAACAGAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA
501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG
551 CCAAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTGTCT
601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG
651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCTGA

```

This encodes a protein having amino acid sequence [SEQ ID 702] (SEQ ID NO: 702):

10

```

1 MKKLSRIVFS TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIRQYSV
51 AEGIAHAQXF XYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
101 KMAGGFSKGK PDGEWVNWYP NGKKSAMVPY KNGLSEGTGX RYYRNGGKES
151 EIQFKQNKAN GVWKQWYADG NIKTEMVMVN DEPAKILTWD ESRLLSELS
201 IHHRXNGVV LEWYEDGSKK XEAVYQDDKL VRKTQWDXDG YLIEP*

```

15

ORF27a (SEQ ID NO: 702) and ORF27-1 (SEQ ID NO: 700) show 94.7% identity in 245 aa overlap:

20	orf27a.pep	10	20	30	40	50	60
		MKKLSRIVFSTVLLGFSAALPAQXYSVYFNQNGKLTATXSAAAYIRQYSVAEGIAHAQXF					
	orf27-1	MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAAYIRQYSVVAGIAHAQDF					
		10	20	30	40	50	60
25	orf27a.pep	70	80	90	100	110	120
		XYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGKPDGEWVNWYP					
	orf27-1	YYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGKPDGEWVNWYP					
		70	80	90	100	110	120
30	orf27a.pep	130	140	150	160	170	180
		NGKKSAMVPYKNGLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN					
	orf27-1	NGKKSAMVPYKNGLSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN					
		130	140	150	160	170	180
35	orf27a.pep	190	200	210	220	230	240
		DEPAKILTWDESGRLLSELSIHHRXNGVVLEWYEDGSKKXEAVYQDDKLVRKTQWDXDG					
	orf27-1	DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG					
		190	200	210	220	230	240
40	orf27a.pep	YLIEPX					
	orf27-1	YLIEPX					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 (SEQ ID NO: 698) shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) (SEQ ID NO: 704) from *N.gonorrhoeae*:

-520-

```

      orf27.pep                                KQWYADXS IKTEMVMVNDEPAKILTWDESG      30
      orf27ng      LSEG TGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDESG      193
5      orf27.pep      RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDGYLIEP      82
      orf27ng      RLLSELSIRHHKRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDGYLIEP      245

```

The complete length ORF27ng nucleotide sequence [<SEQ ID 703>] (SEQ ID NO: 703) is:

```

10      1  ATGAAGAAAT TATCTCGGAT TGTATTTTCA ATCGTACTGT TGGGTTTTTC
      51  GGCCGCTTTG CCGCGCAGAG CCTATTCTGT TTATTTTAAT CAGAACGGGA
      101  AACTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
      151  GCGGCGGGTA TCGCACACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
      201  ATATCCGAA CCTTATATCG TTGCTTCAAC GCAAAATCAA TCTTTTGTGC
      251  CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
15      301  AAAATGGCGG GGGGCTTCAG CAAGGCTAAG CCGGACGGGG AATGGGTCAA
      351  CTGGTATCCG AACGGTAAAA AATCTGCGGT TATGCCTTAT AAAAATGGCT
      401  TGAGTGAGGG TACGGGATAC CGTTATTACC GTAACGGCGG CAAGGAAAGC
      451  GAAATCCAGT TTAAGCAAAA TAAGGCGAAC GCGGTATGGA AGCAATGGTA
20      501  TGCCGATGGA AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
      551  CCAAATTTCT GACTTGGGAT GAAAGCGGCC GATTACTTTC GGAAGTGTCT
      601  ATCCGCCACC ATAAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
      651  TTCTAAAAAG AGCGAGGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
      701  CCCAATGGGA TAAGGATGGT TATTTAATCG AACCTGA

```

25 This encodes a protein having amino acid sequence [<SEQ ID 704>] (SEQ ID NO: 704):

```

      1  MKKLSRIVFS IVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIRQYSV
      51  AAGIAHAQDF YPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
      101  KMAGGFSK GK PDGEWVNWYP NGKSAVMPY KNGLSEGTGY RYYRNGGKES
30      151  EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESRLLSELS
      201  IRHHKRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

```

ORF27ng (SEQ ID NO: 704) and ORF27-1 (SEQ ID NO: 700) show 98.8% identity in 245 aa overlap:

```

35      10      20      30      40      50      60
      orf27-1.pep  MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF
      orf27ng      MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVAAGIAHAQDF
      10      20      30      40      50      60

40      70      80      90      100     110     120
      orf27-1.pep  YPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSK GKPDGEWVNWYP
      orf27ng      YPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSK GKPDGEWVNWYP
      70      80      90      100     110     120

45      130     140     150     160     170     180
      orf27-1.pep  NGKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
      orf27ng      NGKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
      130     140     150     160     170     180

```


		190	200	210	220	230	240
orf27-1.pep		DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG					
orf27ng		DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG					
		190	200	210	220	230	240
orf27-1.pep	YLIEPX						
orf27ng	YLIEPX						

10 Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (SEQ ID NO: 700) (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 (SEQ ID NO: 700) is a surface-exposed protein and a useful immunogen.

Example 84

20 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 705>] (SEQ ID NO: 705):

1 ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
51 TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACKAG CTGTCCGGTT TCTATTGGCA CGCGCATGAg
25 151 ATGATTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGCGGC GTaTCTGGTC
251 GGCTTGACTA TCTTTTGGCT GGCTGCGCGG ATTGCCGCCT TTATCCCGGG
301 TTGGGGTGCG TCGGCAAGCG GCATACTCGG TACGCTGTTT TTCTGGTACG
351 GCGCGGTGTG CATGGCTTTG CCCGTTATCC GTTCGCAGAA TCAACGCAAC
30 401 TATGTTgCCG TGTTCCGCGT GTTCGTCTTG GCGGCACGC ATGCGGCGTT
451 CCACGTCCAG CTGCACAACG GCAACCTAGG CGGACTCTTG AGCGGATTGC
501 AGTCGGGCTT GGTGATG

This corresponds to the amino acid sequence [<SEQ ID 706; ORF47>] (SEQ ID NO: 706; ORF47):

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYTGTHX LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIRSQN QRNYVAVFAL FVLGGTHAAF

151 HVQLHNGNLG GLLSGLQSGL VM

Further work revealed the complete nucleotide sequence [<SEQ ID 707>] (SEQ ID NO: 707):

5
10
15
20
25

```

1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
51  TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTTCTGGTCG
251 GCTTGACTAT CTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCCGGGT
301 TGGGGTGCCT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAACT
401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGCACGCA TCGGGCGTTC
451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
551 TTATTTCGTT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCAGT
601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACTGCCAT
651 GCTGATGGCG CACGCTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCTTTG
701 CGGCAGGTGT GATTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAACCC
751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCCGCTTTCC
851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CATACGGGCA ATCCGATTTA
951 TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
1101 GTGGAAGTAT ATTCTTGGC TGATTCTGCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This corresponds to the amino acid sequence [<SEQ ID 708; ORF47-1>] (SEQ ID NO: 708;

30 ORF47-1):

35

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIRSQN QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIISFFTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAMLMA HGVLAWLSAV FAFAAGVIFT VQVYRWWYKP
251 VLKEPMLWIL FAGYLFITGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

40 Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 (SEQ ID NO: 706) shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) (SEQ ID NO: 710) from strain A of *N. meningitidis*:

45

```

              10      20      30      40      50      60
orf47.pep    MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV
              |||||

```

5	orf47a	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWG	10	20	30	40	50	60
			70	80	90	100	110	120
	orf47.pep	IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC						
	orf47a	IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC						
			70	80	90	100	110	120
10	orf47.pep	MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVM	130	140	150	160	170	
	orf47a	MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI						
			130	140	150	160	170	180
15	orf47a	GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMHGVMPWLSAAFAFAAGVIFT	190	200	210	220	230	240

The complete length ORF47a nucleotide sequence [<SEQ ID 709>] (SEQ ID NO: 709) is:

	1	ATGAAATTTA	CCAAGCACCC	CGTTTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACTGGCG	GCTCTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
20	101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTTGGG	GTTATGCCGG	ACTGGTCGTC	ATCGCCTTCC	TGCTGACCGC
	201	CGTCGCCACT	TGGACGGGGC	AGCCGCCAC	GCGGGCGGC	GTTCTGGTCG
	251	GCTTGACTAT	CTTTTGGCTG	GCTGCGCGGA	TTGCCGCCTT	TATCCCGGGT
	301	TGGGGTGCGT	CGGCAAGCGG	CATACTCGGT	ACGCTGTTTT	TCTGGTACGG
25	351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TTCGCAGAA	CAACGCAATT
	401	ATGTTGCCGT	GTTTCGCGTG	TTCGTCTTGG	GCGGTACGCA	CGCGGCGTTC
	451	CACGTCCAGC	TGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
	501	GTCGGGCTTG	GTGATGGTGT	CGGGTTTAT	CGGTCTGATT	GGTACGCGGA
	551	TTATTTTCGT	TTTTACGTCC	AAACGGTTGA	ATGTGCCGCA	GATTCCAGT
	601	CCGAAATGGG	TGGCGCAGGC	TTCGCTGTGG	CTGCCCATGC	TGACCGCCAT
30	651	GCTGATGGCG	CACGGCGTGA	TGCCTTGGCT	GTCGGCGGCT	TTCGCGTTTG
	701	CGGCAGGTGT	GATTTTTACC	GTGCAGGTGT	ACCGCTGGTG	GTATAAGCCT
	751	GTGTTGAAAG	AGCCGATGCT	GTGGATTCTG	TTTGCCGGCT	ATCTGTTTAC
	801	CGGATTGGGG	CTGATTGCGG	TCGGCGCGTC	TTATTTCAAA	CCCGCTTTCC
	851	TCAATCTGGG	TGTGCATCTG	ATCGGGGTCG	GCGGTATCGG	CGTGCTGACT
35	901	TTGGGCATGA	TGGCGCGTAC	CGCGCTCGGT	CATACGGGCA	ATCCGATTTA
	951	TCCGCCGCC	AAAGCCGTT	CCGTTGCGTT	TTGGCTGATG	ATGGCGGCAA
	1001	CCGCCGTCCG	TATGGTTGCC	GTATTTTCTT	CCGGCACTGC	CTACACGCAC
	1051	AGCATACGCA	CCTCTTCGGT	TTTGTGTTGCA	CTCGCGCTTT	TGGTGATATG
40	1101	GTGGAAGTAT	ATTCTTGGC	TGATTCTGTC	GCGTTCGGAC	GGCAGGCCCC
	1151	GTTGA				

This encodes a protein having amino acid sequence [<SEQ ID 710>] (SEQ ID NO: 710):

	1	MKFTKHPVWA	MAFRPFYSLA	ALYGALSVLL	WGFYGTGTHE	LSGFYWHAHE
45	51	MIWGYAGLVV	IAFLLTAVAT	WTGQPPTRG	VLVGLTIFWL	AARIAAFIPG
	101	WGASASGILG	TLFFWYGAVC	MALPVIRSQN	QRNYVAVFAL	FVLGGTHAAF
	151	HVQLHNGNLG	GLLSGLQSG	VMVSGFIGLI	GTRIIISFFTS	KRLNVPQIPS
	201	PKWVAQASLW	LPMLTAMLM	HGVMPWLSAA	FAFAAGVIFT	VQVYRWYKYP
	251	VLKEPMLWIL	FAGYLTGLG	LIAVGASYFK	PAFLNLGVHL	IGVGGTGVLT
	301	LGM MARTALG	HTGNPIYPPP	KAVPVAFWLM	MAATAVRMVA	VFSSGTAYTH
50	351	SIRTSSVLFA	LALLVYAWKY	IPWLIRPRSD	GRPG*	

ORF47a (SEQ ID NO: 710) and ORF47-1 (SEQ ID NO: 708) show 99.2% identity in 384 aa overlap:

5	orf47a.pep	10 20 30 40 50 60	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV
	orf47-1	10 20 30 40 50 60	
10	orf47a.pep	70 80 90 100 110 120	IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
	orf47-1	70 80 90 100 110 120	
15	orf47a.pep	130 140 150 160 170 180	MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI
	orf47-1	130 140 150 160 170 180	
20	orf47a.pep	190 200 210 220 230 240	GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAGVMPWLSAFAFAAGVIFT
	orf47-1	190 200 210 220 230 240	
25	orf47a.pep	250 260 270 280 290 300	VQVYRWYKPVLPKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
	orf47-1	250 260 270 280 290 300	
30	orf47a.pep	310 320 330 340 350 360	LGMMARTALGHTGNPIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
	orf47-1	310 320 330 340 350 360	
35	orf47a.pep	370 380	LALLVYAWKYIPWLIRPSDGRPGX
	orf47-1	370 380	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 (SEQ ID NO: 706) shows 97.1% identity over 172 aa overlap with a predicted ORF
40 (ORF47ng) (SEQ ID NO: 712) from *N.gonorrhoeae*:

ORF47	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV	60
ORF47ng		60

ORF47 IAFLLTAVATWTGQPPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC 120
 ORF47ng IAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC 120

5 ORF47 MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVM 172
 ORF47ng MALPVIRSQNRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVWGFIGLI 180

The ORF47ng nucleotide sequence [<SEQ ID 711>] (SEQ ID NO: 711) is predicted to encode a protein comprising amino acid sequence [<SEQ ID 712>] (SEQ ID NO: 712):

10 1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
 51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTAFWL AARIAAFIPG
 101 WGAAASGILG TLFFWYGAVC MALPVIRSQN RRNYVAVFAI FVLGGTHAAF
 151 HVQLHNGNLG GLLSGLQSGL VMVWGFIGLI GMKIISFFTS KRLKLPQIPS
 201 PKWVAHASLW LPMLNAILMA HRVMPWLSAA FPFAAGVIFT VQVYAGGITP
 15 251 IEETSCGSVA GICYRLGNSS G

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396 (SEQ ID NO: 1157), accession number e246540):

TM segments in ORF47ng

INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence [<SEQ ID 713>] (SEQ ID NO: 713):

1 ATGAAATTTA CCAAACATCC CGTCTGGGCA ATGGCGTTCC GCCCGTTTTA
 51 TTCACTGGCG GCACTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
 101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
 151 ATGATTGGGG GTTATGCCGG TCTCGTCGTC ATCGCCTTCC TGCTGACCGC
 35 201 CGTCGCCACT TGGACGGGAC AGCCGCCAC GAGGGCGGC GTTCTGGTCG
 251 GCTTGACCGC CTTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
 301 TGGGGTGCGG CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
 351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TtcgCAAAAC CGGCGCAACT
 401 ATGtcgCCGT ATTCGCAATA TTTGTGCTGG GCGGTACGCA TGCGgcgTTC
 40 CACGtccAgc tGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
 501 GTCGGGCTG GTTATGGTGT CGGGCTTTAT CGGCCTGATT GGGATGAGGA
 551 TTATTTCTGT TTTTACGTCC AAACGGTTGA ACGTGCCGCA GATTCCCACT
 601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTACCCATGC TGACCCCAT
 651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
 45 701 CGGCGGGCGT GATTTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
 751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
 801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCTGCCTTCC
 851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT

10

15

20

25

30

35

40

45

250 260 270 280 290 300
 orf47-1.pep VQVYRWYKPV LKEPMLWILFAGYLF TGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
 |||||
 orf47ng-1 VQVYRWYKPV LKEPMLWILFAGYLF TGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
 250 260 270 280 290 300

	310	320	330	340	350	360
orf47-1.pep	LGM	MARTALGHTGNPIY	PPPKAVPVAF	FWLMM	AATAVRMVAV	FSSGTAYTHSIR
orf47nq-1	LGM	MARTALGHTGNSIY	PPPKAVPVAF	FWLMM	AATAVRMVAV	FSSGTAYTHSIR

-527-

	310	320	330	340	350	360
	370	380				
orf47-1.pep	LALLVYAWKYIPWLIRPRSDGRPGX					
5 orf47ng-1	LALLVYAWKYIPWLIRPRSDGRPGX					
	370	380				

Furthermore, ORF47ng-1 (SEQ ID NO: 714) shows significant homology to an ORF (SEQ ID NO: 1157) from *Pseudomonas stutzeri*:

```

10  gnl|PID|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396
    Score = 155 bits (389), Expect = 5e-37
    Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

    Query: 7  PVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59
              P+W +AFRPF+  +LY L++ LW  +TG  GF  WH HEM++G+A  +
15  Sbjet: 14  PIWRLAFRPFFLAGSLYALLAIPLVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

    Query: 60  VIAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119
              V  FLLTAV TWTGQ  G  LVGL A WLAAR+  ++ G  AA  L  LF
    Sbjet: 72  VAGFLLTAVQTTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLLFLVALVW 130

    Query: 120  CMALPVIRSQNRRNYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179
              MA  +  + +RNY  V  + ++ G  +V+  + L
20  Sbjet: 131  MMAQMLWAVRQKRNYPIVVVLSMLGADVILITGLLQGDALQRQGVLAGLWLVAALMAL 190

    Query: 180  IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV---MPWLSAFAFA 234
              IG R+I FFT + L  P  W+  A L  + A+L A GV  P L  F  A
    Sbjet: 191  IGGRVIPFFTQRLGKVDVAVKPWVWLDVALLVGTGVIALHAFGVAMRPQPLLGLLFV-A 249

    Query: 235  AGVIFTVQVYRWYKPVLPKEPMLWILFAGYLFTGLGLIAVGASYF-KPAFXXXXXXXXXXX 293
              GV  +++ RW+ K + K  +LW L  L+  +  +  +F  A
25  Sbjet: 250  IGVGHLRLMRWYDKGIWKVGLLWSLHVAMLWLVAFAFGLALWHFGLLAQSSPSLHALSV 309

    Query: 294  XXXXXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXXFSSTAYTHSIR 353
              M+AR LGHTG  +  P  +  AF L  F S  +
30  Sbjet: 310  GSMSGLILAMIAVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAVPVGGLW 365

    Query: 354  TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384
              ++V + LA  +Y W+Y P L+  R DG PG
    Sbjet: 366  LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 715>] (SEQ ID NO: 715):

1 ..ATGCCGTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAAY CAGAAGyGGT
 51 AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
 101 CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT
 151 GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
 5 201 TGTkGCTTTC GTGATAGGsA GGTTTGyTGG kmksAsyTTG TAyrATwkkG
 251 CCTssCwsTG kAGmGCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTTTTy
 301 TkTtyyCACC GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
 351 CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTGAAATC GTAAATGGCG
 401 GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT
 10 451 TGTTTGGGTT TCTTTGTAGT TGTGTTTAT CTCTTCAGTA ACTTTTTTAG
 501 TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT
 551 ATTCTCCAGC CGCCGAAATC ..

This corresponds to the amino acid sequence [<SEQ ID 716; ORF67>] (SEQ ID NO: 716; ORF67):

1 ..MPSEGSdGXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX
 51 DVETDTGDDT KTXAADXVAF VIGRFxGXXL YXXAXXXAX XWXXXSRGF
 101 XXHRMNLmFN VSVGDARADI GFEFIVEFEI VNGGQAERRN GVEAAVSLMF
 151 CLGFFVvVVY LFSNFFSRRI TFFPFSVTGI ICRYSPAAEI ..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 (SEQ ID NO: 716) shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) (SEQ ID NO: 718) from *N.gonorrhoeae*:

25	orf67.pep	MPSEGSdGXG XGEXEXVAHA QXDFVGFEAG	30
	orf67ng	TNFEIAVLsgMTVRVFYCARPAPVNGRLKMPSEGSdGIGIGESEAVAHaQRGFVGFEAG	146
		90 100 110 120 130 140	
30	orf67.pep	VFQASPVVVT VSGVXXQLGX DVETDTGDDT KTXAADXVAF VIGRFxGXXLYXXAXXXAX	90
	orf67ng	VFQASPVVAVAGVQQA GRD VYAHARHRAEAQAAAVAF LIGVFLRMSVRINRNCCVSI	206
	orf67.pep	XWXXXSRGFXXHRMNLmFN VSVGDARADIGFEFIVEFEI VNGGQAERRNGVEAAVSLMF	150
	orf67ng	TRVGKSTCYFFSRIDAVSDVSGDARTDIGFEFVVEFEI VNGGQAERRNGVECAVFLMF	266
35	orf67.pep	CLGFFVv-----VYLFsNFFSRRI TFF-PFSVTGI ICRYSPAAEI	190
	orf67ng	RLLVFYVKLVAAKSFIILSFQLFYVHGIFIVPFPVTGI IRGDAPAAEVVADRHpgVDGM	326

The ORF67ng nucleotide sequence [<SEQ ID 717>] (SEQ ID NO: 717) is predicted to encode a protein comprising amino acid sequence [<SEQ ID 718>] (SEQ ID NO: 718):

1 MPSETVGSIV NVGVDESvGF SPPFPsIQHF YRFHRIHRIR LFRPPGPMQL
 51 NRHSHGSGNL GRGVWATVLS DKFPCGQVRI PACAGMTNFE IAVLsgMTVR
 101 VFYCARPAPV NGRLKMPSE GSDGIGIGES EAVAHaQRGF VGFEAGvFQA

5 151 SPVVVAVAGV QGQAGRDVYA HARHRAEAQA AAAVAFLIGV FLRMSVRINR
 201 NCCVSITRVG GKSTCYFFSR IDAVSDVSVG DARTDIGFEF VVEFEIVNGG
 251 QAERRNGVEC AVFLMFRLLV FYVKLVAAKS FIILSFQLFY VHGIFIVVPF
 301 PVTGIIRGDA PAAEVVADRH PGVDGMRTDV SEIIAYRAYF VFAWSGWPRI
 351 IVGNAFGGVG *

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 86

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 719>] (SEQ ID NO: 719)

15 1 ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT
 51 TTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
 101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
 151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
 201 CATGTTTCGCC GCCGGACGAA TTTGGGGGCA GARArTCCTA rGGTTCArAC
 251 CTATTGCGsG CATCATGACG CCGrAACGTT ATGAGCAGGT TCAGGAAAAA
 301 TTCGACAAAT ACGGTAACTG GGTCTTATTT GTCGCCCCTT TCCTGCCCGG
 351 TTTGAGAACG GCCGTATTTG TTACAGCCGG TATCAGCCGC AAGGTTTCAT
 401 ACTTGCGTTT TATCATTATG GATGGACTGG CCGCA...

This corresponds to the amino acid sequence [<SEQ ID 720; ORF78>] (SEQ ID NO: 720; ORF78):

25 1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
 51 HIMFAVGMLG VLVGDGIMFA AGRIWGQXXL XFXPIAXIMT PXRYEQVQEK
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAA...

Further work revealed the complete nucleotide sequence [<SEQ ID 721>] (SEQ ID NO: 721):

30 1 ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT
 51 TTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
 101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
 151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
 201 CATGTTTCGCC GCCGGACGAA TTTGGGGGCA GAAAATCCTA AGGTTCAAAC
 251 CTATTGCGCG CATCATGACG CCGAAACGTT ATGAGCAGGT TCAGGAAAAA
 301 TTCGACAAAT ACGGTAACTG GGTCTTATTT GTCGCCCCTT TCCTGCCCGG
 351 TTTGAGAACG GCCGTATTTG TTACAGCCGG TATCAGCCGC AAGGTTTCAT
 401 ACTTGCGTTT TATCATTATG GATGGACTGG CCGCACTGAT TTCCGTCCTT
 451 ATTTGGATTT ATCTGGGCGA ATACGGTGCG CACAAATCG ATTGGCTGAT
 501 GGCGAAAATG CACAGCCTGC AATCGGGTAT TTTTGTTATC TTGGGTATAG
 551 GTGCGACCGT TGTCGCTTGG ATTTGGTGGA AAAAACGCCA ACGTATCCAG
 601 TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA
 651 GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

This corresponds to the amino acid sequence [<SEQ ID 722; ORF78-1>] (SEQ ID NO: 722; ORF78-1):

```

1  MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP
151 IWIYLGEYGA HNIDWLMAKM HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ
201 FYRSKLKEKR AQRKAAKAAK KAAQSKQ*

```

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

10 Homology with the dedA homologue of *H.influenzae* (accession number P45280) (SEQ ID NO: 1158)

ORF78 (SEQ ID NO: 720) and the dedA homologue (SEQ ID NO: 1158) show 58% aa identity in 144aa overlap:

```

15 Orf78: 4  FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61
    FL  FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+      N H+M  V M+GV
DedA:  20  FLIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79

Orf78: 62  LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
    L GD  M+  GRI+G  L F PI  I+T  R   V+EKF +YGN VLFVARFLPGLR
DedA:  80  LAGDSCMYWLGRYGTILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139

20 Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
    +++ +GI+R+VSY+RF+++D  AA
DedA:  140 IYMVSGITRRVSYVRFVLIDFCAA 163

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF78 (SEQ ID NO: 720) shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) (SEQ ID NO: 724) from strain A of *N. meningitidis*:

```

30 orf78.pep      10      20      30      40      50      60
    MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
    |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf78a          10      20      30      40      50      60
    MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG

orf78.pep      70      80      90     100     110     120
    VLVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT
    |||||:||||||||||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf78a          70      80      90     100     110     120
    VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT

```

130 140
 orf78.pep AVFVTAGISRKVSYLRFIIMDGLAA
 |||||
 orf78a AVFVTAGISRKVSYLRFLIMDGLAALISVPVWIYYLGEYGAHNIDWLMAKMHSLSGIFIA
 130 140 150 160 170 180

The complete length ORF78a nucleotide sequence [<SEQ ID 723>] (SEQ ID NO: 723) is:

	1	ATGTTTGCCC	TTTTGGAAGC	CTTTTTTGTC	GAATACGGCT	ATGCGGCCGT
10	51	GTTTTTCGTT	TTGGTCATCT	GCGGTTTCGG	CGTGCCGATT	CCCGAGGATT
	101	TGACCTTGGT	AACAGGCGGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT
	201	CATGTTTCGG	GCCGGACGCA	TCTGGGGGCA	GAAAATCCTC	AAGTTCAAAC
	251	CGATTGCGCG	CATCATGACG	CCGAAACGTT	ACGCACAGGT	TCAGGAAAAA
15	301	TTCGACAAAT	ACGGCAACTG	GGTGTTATTT	GTCGCTCGTT	TCCTGCCCGG
	351	TTTGCGGACT	GCCGTTTTTCG	TTACCGCCGG	CATCAGCCGC	AAAGTATCGT
	401	ATCTGCGCTT	TCTGATTATG	GACGGGCTTG	CCGCGCTGAT	TTCCGTGCCC
	451	GTTTGGAATTT	ACTTGGGCGA	GTACGGCGCG	CACAACATCG	ATTGGCTGAT
	501	GGCGAAAAATG	CACAGCCTGC	AATCCGGCAT	CTTCAATCGCA	TTGGGCGTGC
20	551	TGGCGCGCGC	GCTGGCGTGG	TTCTGGTGGC	GCAAACGCCG	ACATTATCAG
	601	CTTTAGCCGC	CACAATTGAG	CGAAAAACGC	GCCAAACGCA	AGGCGGAAAA
	651	GGCAGCGAAA	AAAGCGGCAC	AGAAGCAGCA	GTAA	

This encodes a protein having amino acid sequence [<SEQ ID 724>] (SEQ ID NO: 724):

25

1	MFALLEAFFV	EYGYAAVFFV	<u>LVICGFGVPI</u>	PEDLTlVtGG	VISGMGYTNP
51	<u>HIMFAVGMLG</u>	VLVGdGIMFA	AGRIWgQKIL	KFKPIARIMT	PKRYAQVQEK
101	FDKYGNWlLF	VARFLPLLRt	AVFVTAGISR	KVSylRFLIM	DGLAALISVP
151	<u>WVIQLGEYGA</u>	HNIDWLMAKM	<u>HSLQSGIFIA</u>	<u>LGVLAAALAW</u>	<u>FWWRKRRHYQ</u>
201	LYRAQLSEKR	AKRKAekAAK	KAAQKQ*		

30 ORF78a (SEQ ID NO: 724) and ORF78-1 (SEQ ID NO: 722) show 89.0% identity in 227 aa overlap:

[illegible]

190 200 210 220

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 (SEQ ID NO: 720) shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) (SEQ ID NO: 726) from *N. gonorrhoeae*:

```

5      orf78.pep      XXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF 137
      orf78ng                      YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF 32
      orf78.pep      IIMDGLAA 145
      orf78ng      LIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLQSGIFIALGVLAALAWFWWRKRR 92
10

```

The ORF78ng nucleotide sequence [<SEQ ID 725>] (SEQ ID NO: 725) is predicted to encode a protein comprising amino acid sequence [<SEQ ID 726>] (SEQ ID NO: 726):

```

15      1  ..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL
      51  GEYGAHNIDW LMAKMHSLQS GIFIALGVLA AALAWFWWRK RRHYQLYRAQ
      101 LSEKRAKRKA EKAAKKAAQK QQ*

```

Further work revealed the complete gonococcal nucleotide sequence [<SEQ ID 727>] (SEQ ID NO: 727):

```

20      1  atgtttgccc tttTggaagc CTTTTTTGTC GAAtacggCt atgcGGCCGT
      51  GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT
      101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
      151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT
25      201 GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC
      251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA
      301 TTCGACAAAT ACGGCAACTG GGTTCGTGTT GTCGCCCGTT TCCTGCCGGG
      351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
      401 ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC
      451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT
30      501 GGCGAAAATG CACAGCCTGC AATCGGGCAT CTTCATCGCA TTGGGCGTGC
      551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
      601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
      651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

```

35 This corresponds to the amino acid sequence [<SEQ ID 728; ORF78ng-1>] (SEQ ID NO: 728; ORF78ng-1):

```

40      1  MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
      51  HIMFAVGMLG VLAGDGVMAF AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
      101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRF LIM DGLAALISVP
      151 VWIYLGEYGA HNIDWLMAM HSLQSGIFIA LGVLAAALAW FWRKRRHYQ
      201 LYRAQLSEKR AKRKA EKAQK*

```

ORF78ng-1 (SEQ ID NO: 728) and ORF78-1 (SEQ ID NO: 722) show 88.1% identity in 227 aa overlap:

```

5      10      20      30      40      50      60
orf78-1.pep MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf78ng-1    MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
      10      20      30      40      50      60

10     70      80      90     100     110     120
orf78-1.pep VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
      ||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf78ng-1    VLAGDGVMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
      70      80      90     100     110     120

15     130     140     150     160     170     180
orf78-1.pep AVFVTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEGYGAHNIDWLMAKMHSLSQSGIFVI
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf78ng-1    AVFVTAGISRKVSYLRFIMDGLAALISVPVWIYLGEGYGAHNIDWLMAKMHSLSQSGIFIA
      130     140     150     160     170     180

20     190     200     210     220
orf78-1.pep LGIGATVVAWIWWKKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX
      ||: ||::||:||:||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
orf78ng-1    LGVLAAALAWFWWRKRRHYQLYRAQLSEKRAKRAEKAAKAAKAAQKQX
      190     200     210     220
```

Furthermore, orf78ng-1 (SEQ ID NO: 728) shows homology to the dedA protein (SEQ ID NO: 1158) from *H.influenzae*:

```

30  sp|P45280|YG29_HAEIN HYPOTHETICAL PROTEIN HI1629 )gi|1073983|pir||D64133 dedA
    protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
    )gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212
    Score = 223 bits (563), Expect = 7e-58
    Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

    Query: 5  LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGVL 62
            L  FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+      N H+M  V M+GVL
    Sbjct: 21  LIGFFTEYGYWAVLFLVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHMLLVSMIGVL 80

35  Query: 63  AGDGVMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122
            AGD  M+  GRI+G KIL+F+PI RI+T +R  V+EKF +YGN VLFVARFLPGLR  +
    Sbjct: 81  AGDSCMYWLGRIYGTKILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140

    Query: 123 FVTAGISRKVSYLRFIMDGLAALISVPVWIYLGEGYGAHNIDWLMAKMHSLSQSGIFIALG 182
            ++ +GI+R+VSY+RF+++D  AA+ISVP+WIIYLG  GA N+DWL  ++  Q  I+I +G
    Sbjct: 141 YMVSGITRRVSYVRFLIDFCAAIISVPIWIYLGELGAKNLDWLHTQIQKGQIVIIYIFIG 200

40  Query: 183 VL 184
            L
    Sbjct: 201 YL 202
```

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 87

- 5 The following partial DNA sequence was identified in *N.meningitidis* [[SEQ ID 729](#)] ([SEQ ID NO: 729](#)):

```

1  ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAAATTCCC GTTACCCTGA AATTAAAAA CGCCAAAGCG CAAACCGTCC
15 401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA C...
```

This corresponds to the amino acid sequence [[SEQ ID 730; ORF79](#)] ([SEQ ID NO: 730; ORF79](#)):

```

20 1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHND EA
51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK QLKKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNH..
```

Further work revealed the complete nucleotide sequence [[SEQ ID 731](#)] ([SEQ ID NO: 731](#)):

```

25 1  ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
201 AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
30 251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAAATTCCC GTTACCCTGA AATTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA
```

- 35 This corresponds to the amino acid sequence [[SEQ ID 732; ORF79-1](#)] ([SEQ ID NO: 732; ORF79-1](#)):

```

40 1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHND EA
51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK QLKKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNHGHH
151 HGEAHQH*
```

Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 (SEQ ID NO: 730) shows 94.6% identity over a 147aa overlap with an ORF (ORF79a)

5 (SEQ ID NO: 734) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf79.pep		MKKLLAAVMMAGLAGAVSAAGVHVEDGWARDTVEGMKIGGAFMKIHNDKQDFLLGGSS					
		: :					
orf79a		MKXLLAAVMMAGLAGAVSAAGIHVEDGWARDTVEGMKMGGAFFMKIHNDKQDFLLGGSS					
10		10	20	30	40	50	60
		70	80	90	100	110	120
orf79.pep		PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP					
		: :					
orf79a		PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP					
15		70	80	90	100	110	120
		130	140				
orf79.pep		VTLKFKNAKAQTVQLEVKIAPMPAMNH					
		:					
orf79a		VTLKFKNAKAQTVQLEVKTAPMSAMDHGHGHGEAHQHX					
20		130	140	150			

The complete length ORF79a nucleotide sequence [<SEQ ID 733>] (SEQ ID NO: 733) is:

	1	ATGAAANAAC	TATTGGCAGC	CGTGATGATG	GCAGGTTTGG	CAGGCGCGGT
	51	TTCCGCCGCC	GGAATCCACG	TTGAGGACGG	CTGGGCGCGC	ACCACCGTCG
25	101	AAGGTATGAA	AATGGGCGGC	GCGTTCATGA	AAATCCACAA	CGACGAAGCC
	151	AAACAAGACT	TTTTGCTCGG	CGGAAGCAGC	CCTGTTGCCG	ACCGCGTCGA
	201	AGTGCATACC	CATATCAATG	ATAACGGTGT	GATGCGGATG	CGCGAAGTCG
	251	AAGGCGCGGT	GCCTTTGGAG	GCGAAATCCG	TTACCGAACT	CAAACCCGGC
30	301	AGCTATCATG	TCATGTTTAT	GGGTNTGAAA	AAACAATTAA	AAGANGGCGA
	351	CAAGATTCCC	GTTACCCTGA	AATTTAAAAA	CGCCAAAGCA	CAAACCGTCC
	401	AACTGGAAGT	CAAAACCGCG	CCGATGTCGG	CAATGGACCA	CGGTCATCAC
	451	CACGGCGAAG	CGCATCAGCA	CTAA		

This encodes a protein having amino acid sequence [<SEQ ID 734>] (SEQ ID NO: 734):

35	1	MKXLLAAVMM	AGLAGAVSAA	GIHVEDGWAR	TTVEGMKMG	AFMKIHNDK
	51	KQDFLLGGSS	PVADRVEVHT	HINDNGVMRM	REVEGGVPLE	AKSVTELKPG
	101	SYHVMFMGXK	QQLKXGDKIP	VTLKFKNAKA	QTVQLEVKTA	PMSAMDHGH
	151	HGEAHQH*				

40 ORF79a (SEQ ID NO: 734) and ORF79-1 (SEQ ID NO: 732) show 94.9% identity in 157 aa overlap:

-536-

		10	20	30	40	50	60
	orf79a.pep	MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGAFMKIHNDEAKQDFLLGGSS					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 (SEQ ID NO: 730) shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) (SEQ ID NO: 736) from *N.gonorrhoeae*:

20	orf79.pep	FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS	101
	orf79ng	INDNGVMRMREVKGVPLEAKSVTELKPGS	30
	orf79.pep	YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMPAMNH	147
25	orf79ng	YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH	86

An ORF79ng nucleotide sequence [<SEQ ID 735>] (SEQ ID NO: 735) was predicted to encode a protein comprising amino acid sequence [<SEQ ID 736>] (SEQ ID NO: 736):

30	1	..INDNGVMRMREVKGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPV
	51	TLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH*

Further work revealed the complete gonococcal DNA sequence [<SEQ ID 737>] (SEQ ID NO: 737):

35	1	ATGAAAAAAT	TATTGGCAGC	CGTGATGATG	GCAGGTTTGG	CAGGCGCGGT
	51	TTccgccgCc	GGagTccAtG	TCGAggACGG	CTGGGCGCGc	accactGtccg
	101	aaggtATgaa	aatggGCGGC	GCgttCATga	aaATCCACAA	CGACGaaGcc
	151	atacaaGACT	ttgtgcTCgg	CGGaagcatg	cccgttgccg	accgcGTCTGA
	201	AGTGCAtaca	cacATCAACG	ACAACGGCGT	GATGCGTATG	CGCGAAGTCA
	251	AAGGCGGCGT	GCCTTTGGAG	GCGAAATCCG	TTACCGAACT	CAAACCCGGC
	301	AGCTATCACG	TGATGTTTAT	GGGTTTGAAA	AAACAACCTGA	AAGAGGGCGA
40	351	CAAGATTCCC	GTTACCCTGA	AATTTAAAAA	CGCCAAAGCG	CAAACCGTCC
	401	AACTGGAAGT	CAAAACCGCG	CCGATGTCGG	CAATGAACCA	CGGTCATCAC
	451	CACGGCGAAG	CGCATCAGCA	CTAA		

This corresponds to the amino acid sequence [<SEQ ID 738; ORF79ng-1>] (SEQ ID NO: 738; ORF79ng-1):

```

1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMGG AFMKIHNDDEA
51 IQDFVLGGSM PVADRVEVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNKA QTVQLEVKTA PMSAMNHGHH
151 HGEAHQH*

```

ORF79ng-1 (SEQ ID NO: 738) and ORF79-1 (SEQ ID NO: 732) show 95.5% identity in 157 aa overlap:

```

10      10      20      30      40      50      60
orf79-1.pep MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHNDDEAKQDFLLGGSS
orf79ng-1    MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKMGGA FMKIHNDDEAIQDFVLGGSM
           10      20      30      40      50      60
15      70      80      90      100     110     120
orf79-1.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
orf79ng-1    PVADRVEVHTHINDNGVMRMREV KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
           70      80      90      100     110     120
20      130     140     150
orf79-1.pep VTLKFKNKAQTVQLEVKIAPMPAMNHGHHHGEAHQH*
orf79ng-1    VTLKFKNKAQTVQLEVKTAPMSAMNHGHHHGEAHQH*
           130     140     150

```

Furthermore, ORF79ng-1 (SEQ ID NO: 738) shows significant homology to a protein (SEQ ID NO: 1159) from *Aquifex aeolicus*:

```

30 gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
   Score = 63.6 bits (152), Expect = 6e-10
   Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

Query: 24  VEDGWAR TTVEGMKMGGA FMKIHNDDEAIQDFVLGGSM PVADRVEVHTHINDNGVMRMREV 83
          V+  W      G      M I N+  D+++G  +A RVE+H  + +N V +M
Sbjct: 27  VKHPWVMEPPPGPNTTMMGMIIIVNEGDEPDYLIGAKTDIAQRVELHKTVIENDVAKMVPQ 86

Query: 84  KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNKAQTVQLEV 137
          + + + K  E K  YHVM +GLKK++KEGDK+ V L F+ +  TV+  V
Sbjct: 87  ER-IEIPKKGKVEFKHHGYHVMIIIGLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (SEQ ID NO: 732) (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-

PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 (SEQ ID NO: 732) is a surface-exposed protein, and that it is a useful immunogen.

5 Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 739>] (SEQ ID NO: 739):

```

1  ATGACGGTAA CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
51  ATATCTGATT ACGGGCATT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
10  101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
151  151 CCGAAGCAAT GCGCGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
201  201 GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGCCG
251  251 CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
301  301 CGGATTCCGG TTGTGAAATC CATCTATTCT AGTGTGAAAA AAGTATCCGA
15  351 ATAcgTGCTG TCCGACAGCA GCCGTTTCGT TAAACGCCC GTACTCGTGC
401  401 CGTTTCCCCA GCCCGGTATT TGGACGATyG CTTTCGTGTC AGGGCAGGTG
451  451 TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTCCGT
501  501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
551  551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AsCATTGAAA
20  601 TATGTGATTT CGCTGGGTAT GGTCAATCCCT GACGACCTGC CCGTCAAAAC
651  651 ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTGCCC GAACAACAAT
701  701 AA

```

This corresponds to the amino acid sequence [<SEQ ID 740; ORF98>] (SEQ ID NO: 740; ORF98):

```

1  MTVTAAEGGK AAKALKKLYI TGILVWLPIA VTVWVSYIV SASDQLVNLL
51  PKQWRPQYVL GFNIPGLGVI VAIAVLFVTG LFAANVLGRQ ILAAWDSLLG
101  RIPVVKSIYS SVKKVSEYVL SDSSRSFKTP VLVPFPQPGI WTIAFVSGQV
151  SNAVKALPX DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEXLK
30  201 YVISLGMVIP DDLPVKTLAX PMPSEKADLP EQQ*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 741>] (SEQ ID NO: 741):

```

1  ATGACGGAAC nTGCGGCCGA AGGCGGCAAA GCTGCCAArG CGTTAAAAAA
35  51  ATATCTGATT ACGGGCATT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
101  101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
151  151 CCGAAGCAAT GCGCGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
201  201 GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGCCG
251  251 CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
301  301 CGGATTCCGG TTGTGAAATC CATCTATTCT AGTGTGAAAA AAGTATCCGA
40  351 ATCGCTGCTG TCCGACAGCA GCCGTTTCGT TAAACGCCC GTACTCGTGC
401  401 CGTTTCCCCA GCCCGGTATT TGGACGATTG CTTTCGTGTC AGGGCAGGTG
451  451 TCGAATGCGG TTAAGGCCGC ATTGCCGAAG GACGGCGATT ATCTTCCGT
501  501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
551  551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCATTGAAA
45  601 TATGTGATTT CGCTGGGTAT GGTCAATCCCT GACGACCTGC CCGTCAAAAC

```

651 ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTGCCCC GAACAACAAT
701 AA

This corresponds to the amino acid sequence [<SEQ ID 742; ORF98-1>] (SEQ ID NO: 742;

5 ORF98-1):

1 MTEXAAEGGK AAKALKKYL I TGILVWLPIA VTVWVVSIV SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGVI VAIAVLFVTG LFAANVLGRQ ILAAWDSLLG
101 RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPQPGI WTIAFVSGQV
151 SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
201 YVISLGMVIP DDLPVKTLAG PMPSEKADLP EQQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 (SEQ ID NO: 740) shows 96.1% identity over a 233aa overlap with an ORF (ORF98a)

15 (SEQ ID NO: 744) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf98.pep		MTVTAAEGGKAAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL					
orf98a		MTEPAAEGGKAAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL					
		10	20	30	40	50	60
orf98.pep		GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL					
orf98a		GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVXSLL					
		70	80	90	100	110	120
orf98.pep		SDSSRSFKTPVLVFPQPGIWTIAFVSGQVSNVKAALPXDGDYLSVYVPTTPNPTGGYY					
orf98a		SDSSRSFKTPVLVFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
		130	140	150	160	170	180
orf98.pep		IMVKKSDVRELDMSVDEXLKYVISLGMVIPDDLVPKTLAXPMPSEKADLPEQQX					
orf98a		IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPSEKADLPEQQX					
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence [<SEQ ID 743>] (SEQ ID NO: 743) is:

1 ATGACGGAAC CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
51 ATATCTGATT ACGGGCATT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
151 CCGAAGCAAT GCGCGCCGCA ATATGTTTGG GGGTTTAATA TCCCGGGGCT
201 GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTATTGCGG
251 CAAACGTATT GGGCCGCGAG ATTCTTGCCG CGTGGGACAG CTTGTTGGGG
301 CGGATTCCGG TTGTGAAGTC CATCTATTTC AGTGTGAAAA AAGTATCCGA

-540-

5
351 NTCGTTGCTG TCCGACAGCA GCCGTTTCGTT TAAAACACCA GTACTCGTGC
401 CGTTTCCCCA ATCGGGTATT TGGACAATCG CATTCGTGTC CGGTCAGGTG
451 TCGAATGCGG TTAAGGCCGC ATTGCCGAAG GACGGCGATT ATCTTTCCGT
501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGACGA AGCGTTGAAA
601 TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
651 ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTGCCC GAACAACAAT
701 AA

10 This encodes a protein having amino acid sequence [<SEQ ID 744>] (SEQ ID NO: 744):

15
1 MTEPAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGVI VAIIVLFVTG LFAANVLGRQ ILAAWDSLLG
101 RIPVVKSIYS SVKKVSXSL SDSSRSFKTP VLVFPQSGI WTIAFVSGQV
151 SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
201 YVISLGMVIP DDLPVKTLAG PMPSEKADLP EQQ*

ORF98a (SEQ ID NO: 744) and ORF98-1 (SEQ ID NO: 742) show 98.7% identity in 233 aa overlap:

20	orf98a.pep	10	20	30	40	50	60
		MTEPAAEGGKAAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLLPKQWRPQYVL					
	orf98-1	MTEXAAEGGKAAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLLPKQWRPQYVL					
		10	20	30	40	50	60
25	orf98a.pep	70	80	90	100	110	120
		GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSXSL					
	orf98-1	GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL					
		70	80	90	100	110	120
30	orf98a.pep	130	140	150	160	170	180
		SDSSRSFKTPVLVFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
	orf98-1	SDSSRSFKTPVLVFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
		130	140	150	160	170	180
35	orf98a.pep	190	200	210	220	230	
		IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAGPMPSEKADLPEQQX					
	orf98-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAGPMPSEKADLPEQQX					
		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF98 (SEQ ID NO: 740) shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) (SEQ ID NO: 746) from *N.gonorrhoeae*:

45	orf98.pep	10	20	30	40	50	60	60
		MTVTAAEGGKAAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLLPKQWRPQYVL						
	orf98ng	MTEPAAEGGKAAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLLPKQWRPQYVL						60

-541-

```

orf98.pep    GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL 120
              |||||
orf98ng      GFNIPGLGVIVAIIVLFVTGLFAANVLGRQILAAWDSLLXRIPVVKSIYSSVKKVSESL 120

orf98.pep    SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVSNVKAALPXDGDYLSVYVPTTPNPTGGYY 180
              |||||
orf98ng      SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPQDGDYLSVYVPTTPNPTGGYY 180

orf98.pep    IMVKKSDVRELDMSVDEXLKYVISLGMVIPDDLVPVKTLAXPMPSEKADLPEQQ 233
              |||||
orf98ng      IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPPEKAELPEQQ 233

```

The complete length ORF98ng nucleotide sequence [[SEQ ID 745](#)] ([SEQ ID NO: 745](#)) is predicted to encode a protein having amino acid sequence [[SEQ ID 746](#)] ([SEQ ID NO: 746](#)):

```

1  MTEPAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGVI VAIIVLFVTG LFAANVLGRQ ILAAWDSLLX
101 RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPFQSGI WTIAFVSGQV
151 SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
201 YVISLGMVIP DDLVPVKTLG PMPPEKAELP EQQ*

```

Further work revealed the complete nucleotide sequence [[SEQ ID 747](#)] ([SEQ ID NO: 747](#)):

```

1  ATGACGGAAC CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
51 ATATCTGATT ACAGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ACCAGCTTGT CAACCTGCTG
151 CCGAAGCAAT GCGGCGCCGA ATATGTTTTG GGGTTTAATA TCCCCGGGCT
201 CCGCGTTATT GTTGCCATTG CCGTATTGTT TGTAACCGGA TTATTTGCCG
251 CAAACGTGTT GGGCCGCGAG ATTCTTGCCG CGTGGGACAG CCTGTTgggg
301 cggatTCCGG TTGTCAAATC CATCTATTCT AGTGTGAAAA AAGTATCCGA
351 ATCGCTGCTG TCCGACAGCA GCCGTTCGTT TAAAACGCCG GTACTCGTGC
401 CGTTTCCCCA ATCGGGTATT TGGACAATCG CATTCGTGTC CGGTCAGGTG
451 TCGAATGCGG TTAAGGCCGC ATTGCCGAG GATGGCGATT ATCTTTCCGT
501 GTATGTCCCG ACCACGCCCA ACCCGACCGG CGGTACTAT ATTATGGTAA
551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGACGCA AGCGTTGAAA
601 TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
651 ATTGGCAGGA CCTATGCCGC CTGAAAAGGC GGAGTTGCC GAACAACAAT
701 AA

```

This corresponds to the amino acid sequence [[SEQ ID 748; ORF98ng-1](#)] ([SEQ ID NO: 748; ORF98ng-1](#)):

```

1  MTEPAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL
51 PKQWRPQYVL GFNIPGLGVI VAIIVLFVTG LFAANVLGRQ ILAAWDSLLG
101 RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPFQSGI WTIAFVSGQV
151 SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
201 YVISLGMVIP DDLVPVKTLG PMPPEKAELP EQQ*

```

ORF98ng-1 ([SEQ ID NO: 748](#)) and ORF98-1 ([SEQ ID NO: 742](#)) show 97.9% identity in 233 aa overlap:

```

              10      20      30      40      50      60
orf98-1.pep  MTEXAAEGGKAALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL

```


1051 GGCAAAAGGC TACCTTGAAG CGAGCATTGC ATTAAAGCCG AGTATTTCCG
 1101 CGCGTTTGGT TCTAACAAAG GTTTTCGACG AAATCGGAGA ACCGCAGAAG
 1151 GCGGAGGCGC AC...

5 This corresponds to the amino acid sequence [<SEQ ID 750; ORF100>] (SEQ ID NO: 750; ORF100):

1 MKTVVWIVVL FAAAVGLALA SGIYTG DVYI VLGQTMLRIN LHAFVLGSLI
 51 AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GXKXXLALNK AGLAYFEGRF
 101 EKAELEASRV LVNKGVRDNR TLALMLXAHA AGQMENIXXR DRYLAEIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAKMNANLT RLVRLXIRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKVQHYP XNRRPELLEA
 301 FVESVRFLGE REQQAIDFA DAWLKEQPDN ALLLMYLGR L AFGRKLWGKA
 351 KGYLEASIAL KPSISARLVL TKVFDEIGEP QKAEAH...

Further work revealed the complete nucleotide sequence [<SEQ ID 751>] (SEQ ID NO: 751):

1 ATGAAAACGG TAGTCTGGAT TGTCGTCCTG TTGCGCGCCG CCGTCGGACT
 51 GGCGCTGGCT TCGGGCATTT ACACCGGCGA CGTGTATATC GTACTCGGAC
 101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTTAGG TTCGCTGATT
 151 GCCGTCGTGG TGTGGTATTT CTTGTTTAAA TTCATTATCG GCGTACTCAA
 201 TATCCCGGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTTT
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA
 401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
 451 CCGGAAAAAC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT
 501 GAACCGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCCA
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
 601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAACTTTC
 651 CAAGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
 701 CATACGCCG CCAGCTGGCG GATGCTGCCG ATGCCGCCG TTTGAAAACC
 751 TGCCTGAAGC GGATCCCCGA CAGCCTCAA AACGGGAAT TGAGCGTATC
 801 GGTTCGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
 851 GGGTCAAACA CATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
 901 TTTGTCGAAA GCGTGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
 951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
 1001 TGATGTATCT CGGTCGGCTC GCCTACGGCC GCAAACTTTG GGGCAAGGCA
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG
 1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACCG CAGAAGGCGG
 1151 AGGCGCAGCG CAACTTGTTT TTGGAAGCCG TCTCCGATGA CGAACGTAC
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence [<SEQ ID 752; ORF100-1>] (SEQ ID NO: 752; ORF100-1):

1 MKTVVWIVVL FAAAVGLALA SGIYTG DVYI VLGQTMLRIN LHAFVLGSLI
 51 AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF
 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAKMNANLT RLVRLQLRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKVQHYP HNRRPELLEA
 301 FVESVRFLGE REQQAIDFA DAWLKEQPDN ALLLMYLGR L AYGRKLWGKA
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRLV LEAVSDDERH

401 AALEQHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 5 ORF100 (SEQ ID NO: 750) shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) (SEQ ID NO: 754) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf100.pep	MKT	VVWIVLFAAAVGLALASGIY	TGDVYIVLGQ	TMLRINLHAFVLGSLIAVVVWYFLFK		
10	orf100a	MKT	VVWIVLFAAAXGLALASGIX	TGDVYIVLGQ	TMLRINLHAFVLGSLIAVVVWYFLFK		
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf100.pep	FIIGVLNIPEKMQRFGSARKG	XXLALNKAGLAYFEGRFEKAELEASRVLVNKVGRDNR				
15	orf100a	FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGNKEAGDNR					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf100.pep	TLALMLXAHAAAGQMENIXRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
20	orf100a	TLALMLGAHAAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
25	orf100a	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf100.pep	DAADAAALKTKLKRI	PD	SLKNGELSVSVAEKYERLGLYADAVKWVKQHYPXNRRPELLEA			
30	orf100a	DAADAAALKTKLKRI	PD	SLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRPELLEA			
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAFGRKLWGKAKGYLEASIAL					
35	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXYLGRLAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
		370	380				
	orf100.pep	KPSISARLVLT	KVFDEIGEPQKAEAH				
40	orf100a	KPSISARLVLA	KVFDETGEPQKAEAQRNVLASVAEENRPSAETHX				
		370	380	390	400		

The complete length ORF100a nucleotide sequence [<SEQ ID 753>] (SEQ ID NO: 753) is:

1 ATGAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CNNTCGGGCT

-545-

5
10
15
20
25

```

51  GGCATTGGCG TCGGGCATTN ACACCGGCGA CGTGTATATC GTACTCGGAC
101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTTAGG TTCGCTGATT
151 GCCGTCGTGG TGTGGTATTT CCTGTTCAAA TTCATCATCG GCGTACTCAA
201 TANCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG
251 CCGCGCTTGC TTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT
301 GAAAAGCGCG AACTTGAAGC CTCGCGCGTA TTGGGAAACA AAGAGGCGGG
351 GGATAACCGG ACTTTGGCAT TGATGTTGGG CGCACATGCC GCCGGGCAGA
401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
451 CCGGAAAAGC AGCAGCTTTC CCGTTATCTT TTGTGCGCGG AATCGGCGTT
501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAANTTTC
651 CAAGGCGGGC GCGTNGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
701 CATACCGCCG CCAGCTGNCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATCCCCGA CAGCCTCAAA AACGGGGAAT TGAGCGTATC
801 GGTTGCGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GACCCGAACT TTTGGAAGCN
901 TTTGTCGAAA GCGTGCCTTT TTTGGGCGAA CGCGATCAGC AGAAAGCCAT
951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAT GCGCTTCTGC
1001 TGANGTATCT CGGTCGGCTC GCCTACGGCC GCAAACCTTG GGGCAAGGCA
1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG
1101 TTTGGTCTG GCAAAGGTTT TTGACGAAAC CGGAGAACCG CAGAAGGCGG
1151 AGGCGCAGCG CAACTTGTTT TTGGCAAGCG TTGCCGAGGA AAACCGNCCT
1201 TCCGCCGAAA CCCATTGA

```

This encodes a protein having amino acid sequence [SEQ ID 754] (SEQ ID NO: 754):

30
35

```

1  MKTVVWIVVL FAAAXGLALA SGIXTGDVYI VLGQTMLRIN LHAFVLGSLI
51  AVVVWYFLFK FIIGVLNXPE KMQRFGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNAULT RLVRLQLRYA
201 FDRGDALQVL AKTEKXSKAG AXGKSEMERY QNWAYRRQLX DAADAAALKT
251 CLKRIPDSLK NGELSVSAE KYERLGLYAD AVKWKQHYH HNRRLPELLEA
301 FVESVRFGE RDQQAIDFA DAWLKEQPDN ALLLXYLGRL AYGRKLWGKA
351 KGYLEASIAL KPSISARLVL AKVFDETGEF QKAEAQRLNV LASVAEENRP
401 SAETH*

```

ORF100a (SEQ ID NO: 754) and ORF100-1 (SEQ ID NO: 752) show 95.1% identity in 406 aa overlap:

40
45
50

```

      10      20      30      40      50      60
orf100a.pep  MKTVVWIVVLFFAAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
              |||||
orf100-1     MKTVVWIVVLFFAAAVGLALASGIYTGVDYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK
              |||||
      10      20      30      40      50      60

      70      80      90     100     110     120
orf100a.pep  FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGNKEAGDNR
              |||||
orf100-1     FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR
              |||||
      70      80      90     100     110     120

      130     140     150     160     170     180
orf100a.pep  TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAAEANLH
              |||||
orf100-1     TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAAEANLH
              |||||
      130     140     150     160     170     180

```

-546-

5	orf100a.pep	190	200	210	220	230	240
	orf100-1	190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 (SEQ ID NO: 750) shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) (SEQ ID NO: 756) from *N.gonorrhoeae*:

25	orf100.pep	MKTVVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
30	orf100.pep	FIIGVLNIPEKMQRFGSARKGXKXXLALNKAGLAYFEGRFEKAELEASRVLVNKVGRDNR	120
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGKKEAGDNR	120
35	orf100.pep	TLALMLXHAHAAGQMENIXRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
	orf100ng	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
40	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	240
	orf100ng	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA	240
	orf100.pep	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPXNRRPELLEA	300
	orf100ng	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRRPELLEA	300
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAFGRKLWGKAKGYLEASIAL	360
	orf100ng	FVESVRFLGEREQQKAIDFADSWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	360
	orf100.pep	KPSISARLVLTKEVDEIGEPQKAEAH	386
	orf100ng	KPSIPARLVLAKEVDETAQSQKAEQARNLVLASVAGENRPSAETR	405

The complete length ORF100ng nucleotide sequence [<SEQ ID 755>] (SEQ ID NO: 755) is:

```

1  ATGAAAACGG TAGTCTGGAT TGTGTGCTCTG TTTGCCGCGG CCGTCGGACT
5  51  GCGGCTGGCT TCGGGCATT TACACCGCGA CGTGTATATC GTACTCGGAC
101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTAGG TTCGCTGATT
151 GCCGTCGTGG TGTGGTATTT CCTGTTTAAA TTCATCATCG GCGTACTCAA
201 TATCCCCGAA AATATGCGGC GTTCCGGTTC GCGCGGAAA GCGCCGAAGG
251 CCGCGCTTGC CTTGAATAAG GCGGGTTTGG CGTATTTCGA AGGGCGTTT
301 GAAAAGCGCG AACTCGAAGC CTCTCGAGTG TTGGGCAACA AAGAGCCGG
10 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCG GCAGGACAGA
401 TGGAAAATAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAACTG
451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGCGGTT
501 AAACCGGCGC GATTACGAAG CCGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
15 601 TTCGATCGGG GCGATGCGTT GCAGGTTCTG GCAAAAaccG AAAAACTTTC
651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGTAT CAAAATTGGG
701 CATAACGCGC CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACGGGAAT TGagcGTATC
801 GGTTCGCGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
20 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
901 TTTGTGAAA GCGTGCGCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
951 CGATTTTGCC GATTCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
1001 TGATGTATCT CGGCCGGCTC GCCTACGGCC GCAAACCTTG GGGTAAGGCA
1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCCGGCGCG
25 1101 TTTGGTGTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAAGCCG
1151 AAGCACAGCG CAACTTGGTT TTGGCAAGCG TTGCCGGGGA AAACGCCCT
1201 TCCGCCGAAA CCCGTTGA

```

This encodes a protein having amino acid sequence [<SEQ ID 756>] (SEQ ID NO: 756):

```

30 1  MKTVVWIVVL FAAAVGLALA SGIYTG DVYI VLQQTMLRIN LHAFVLGSLI
51  AVVVWYFLFK FIIGVLNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAMNANLT RLVRLQLRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAALKT
35 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYP HNRRPELLEA
301 FVESVRFLGE REQKAIDFA DSWLKEQPDN ALLLMYLGRL AYGRKLWGKA
351 KGYLEASIAL KPSIPARLVL AKVFDETAQS QKAEAQRLV LASVAGENRP
401 SAETR*

```

40 ORF100ng (SEQ ID NO: 756) and ORF100-1 (SEQ ID NO: 752) show 95.3% identity in 402 aa overlap:

```

10      20      30      40      50      60
orf100-1.pep MKTVVWIVVLFAAAVGLALASGIYTG DVYI VLQQTMLRINLHAFVLGSLIAVVVWYFLFK
45  orf100ng   MKTVVWIVVLFAAAVGLALASGIYTG DVYI VLQQTMLRINLHAFVLGSLIAVVVWYFLFK
      10      20      30      40      50      60
70      80      90      100     110     120
orf100-1.pep FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR
50  orf100ng   FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR
      70      80      90      100     110     120

```

-548-

		130	140	150	160	170	180
	orf100-1.pep	TLALMLGAHAAGQ	MENIELRDRYLAE	IAKLPEKQQLSRY	LLAESALNRRDY	EAAEANLH	
5	orf100ng	TLALMLGAHAAGQ	MENIELRDRYLAE	IAKLPEKQQLSRY	LLAESALNRRDY	EAAEANLH	
		130	140	150	160	170	180
	orf100-1.pep	AAAKMNANLTRLV	RQLRYAFDRGDAL	QVLAKTEKLSKAG	ALGKSEMERYQN	WAYRRQLA	
10	orf100ng	AAAKMNANLTRLV	RQLRYAFDRGDAL	QVLAKTEKLSKAG	ALGKSEMERYQN	WAYRRQMA	
		190	200	210	220	230	240
	orf100-1.pep	DAADAAALKTKL	KRIPDSLKNGEL	SVSVAEKYERLGL	YADAVKWVKQHYP	HNRRPELLEA	
15	orf100ng	DAADAAALKTKL	KRIPDSLKNGEL	SVSVAEKYERLGL	YADAVKWVKQHYP	HNRRPELLEA	
		250	260	270	280	290	300
	orf100-1.pep	FVESVRFLGEREQ	QKAIDFADAWLKE	QPDNALLMYLGRL	AYGRKLWGKAKGY	LEASIAL	
20	orf100ng	FVESVRFLGEREQ	QKAIDFADSWLKE	QPDNALLMYLGRL	AYGRKLWGKAKGY	LEASIAL	
		310	320	330	340	350	360
	orf100-1.pep	KPSISARLVLA	KVFDEIGEPQKAE	AQRNLVLEAVSD	DERHAALEQHSX		
25	orf100n	KPSIPARLVLA	KVFDETAQSQKAE	AQRNLVLASVAG	ENRPSAETRX		
		370	380	390	400		

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 757>] (SEQ ID NO: 757)

35	1	ATGATGTTTT	CTTGTTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTTCTGTG
	51	GTTTGCAGGG	CTGTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
	101	TTGATGTGCC	GCGCGGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
	151	GTGCGGCTGT	ACCGTTTTAT	GTCGCCGTTG	GGCTTCGGCG	CGGTCGTGTT
	201	CGGCGCGGCG	ATACCGTTTG	CCGCCGCTG	GTGGGGCAGC	GGCTGGGTAC
	251	ACGTCAAAC	TGTTTTGGGC	TTGATGCTCT	TGGCTTACCA	GTTGTATTGC
40	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
	351	CTGGTACCGC	GTGTTCAACG	AAATCCCCGT	GCTGCTGATG	GTTGCCGCGC
	401	TGTATsTGGT	CGTGTTCAAA	CCGTTTTGA		

This corresponds to the amino acid sequence [<SEQ ID 758; ORF102>] (SEQ ID NO: 758; ORF102):

```

1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYXVVF K PF*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 759>] (SEQ ID NO: 759):

```

1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCAATTCGTG
51  GTTTCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGTCTGTC GGGCATGGCG
151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CCGTCTGTT
201 CGGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGCGAGC GGCTGGGTAC
251 ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
301 GGCGTGCTGC TGCCCGGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
15 351 CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
401 TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This corresponds to the amino acid sequence [<SEQ ID 760; ORF102-1>] (SEQ ID NO: 760; ORF102-1):

```

1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
20 51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYL VVF K PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number
25 AE000647) (SEQ ID NO: 1160)

ORF102 (SEQ ID NO: 758) and HP1484 (SEQ ID NO: 1160) show 33% aa identity in 143aa overlap:

```

30 orf102 3  FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
    F W K FH+ VISW A LFYLPR+ FV A + V++ +LY F++
    HP1484 8  FLWVKAFHVIAVISWMAALFYLPRL FVYHAENAHKKEFVG VVQIQEK--KLYSFIASPAM 65

    orf102 63  GAVVFGAAIPFAAG---WWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY 119
    G + + + GW+H KL L ++LLAY YC +R + + R+Y
    HP1484 66  GFTLITGILMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRNARFY 125

35 orf102 120 RVFNEIPXXXXXXXXXXXXXFKPF 142
    RVFNE P KPF
    HP1484 126 RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 (SEQ ID NO: 758) shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) (SEQ ID NO: 762) from strain A of *N. meningitidis*:

```

5      10      20      30      40      50      60
orf102.pep MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
          |||||
orf102a    MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
          10      20      30      40      50      60

10     70     80     90     100    110    120
orf102.pep GFGAVVFGAAIPFAAGWWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR
          |||||
orf102a    GFGAVVFGAAIPFAAGWWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR
          70     80     90     100    110    120

15     130    140
orf102.pep VFNEIPVLLMVAALYXVVF K PFX
          |||||
orf102a    VFNEIPVLLMVAALYL VVF K PFX
          130    140

```

The complete length ORF102a nucleotide sequence [<SEQ ID 761>] (SEQ ID NO: 761) is:

```

20      1  ATGATGTTTT CTTGTTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
      51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTG GGTTCGCGC CGGTCGTGTT
     201  CGGCGCGGCG ATACCGTTTG CCGCCGCTG GTGGGGCAGC GGCTGGGTAC
     25  251  ACGTCAAAC TGTTTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTATGC
     301  GGGCTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

30 This encodes a protein having amino acid sequence [<SEQ ID 762>] (SEQ ID NO: 762):

```

      1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
     51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHV KLC LGLM L LAYQLY
    101  GVL LRRFQDY SNAF SHRWYR VFNEIPVLLM VAALYL VVFK PF*

```

35 ORF102a (SEQ ID NO: 762) and ORF102-1 (SEQ ID NO: 760) show complete identity in 142 aa overlap:

```

40     10     20     30     40     50     60
orf102a.pep MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
          |||||
orf102-1    MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
          10     20     30     40     50     60

45     70     80     90     100    110    120
orf102a.pep GFGAVVFGAAIPFAAGWWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR
          |||||
orf102-1    GFGAVVFGAAIPFAAGWWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR
          70     80     90     100    110    120

```

-551-

5

orf102a.pep	VFNEIPVLLMVAALYL	130	140	VVFKPFX
orf102-1	VFNEIPVLLMVAALYL	130	140	VVFKPFX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 (SEQ ID NO: 758) shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) (SEQ ID NO: 764) from *N. gonorrhoeae*:

10

orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	60
orf102ng	MMFSWFKLFHLFFVISWFAGLFYLPRI	60
orf102.pep	GFGAVVFGAAIPFAAGWWSGWVHV	120
orf102ng	GFGAVVFGAAIPFAAGWWSGWVHV	120
orf102.pep	VFNEIPVLLMVAALYXVVKPF	142
orf102ng	VFNEIPVLLMVAALYL	142

15

The complete length ORF102ng nucleotide sequence [<SEQ ID 763>] (SEQ ID NO: 763) is:

20

1	ATGATGTTTT	CTTGGTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTTCTGTG
51	GTTTGCAGGG	CTGTTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
101	TTGATGCGCC	GCGCGGCAAT	CCCAGATATG	TGCGCCTGTC	GGGGATGGCG
151	GTGCGGTTGT	ACCGTTTTAT	GTCGCCTTTG	GGTTTCGGCG	CGGTCGTGTT
201	CGGCGCGGCG	ATACCGTTTG	CCGCcggcgc	GTGGGGCagc	ggctggGTTC
251	ACGTCAAAC	GTGTTTGGGC	TTGATGCTCT	TGGCTTATCA	GTTGTATTGC
301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
351	CTGGTACCGC	GTGTTCAAcg	aAATCCCCGT	GCTGCTGATG	GTTGCCGCGC
401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

30 This encodes a protein having amino acid sequence [<SEQ ID 764>] (SEQ ID NO: 764):

1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVN MAMIDAPRGN	PEYVRLSGMA
51	VRLYRFMSPL	GFGAVVFGAA	IPFAAGRWGS	GWVHVKLCLG	LMLLAYQLYC
101	GVLLRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYL	VVFK PF*

35 ORF102ng (SEQ ID NO: 764) and ORF102-1 (SEQ ID NO: 760) show 98.6% identity in 142 aa overlap:

40

orf102-1.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	10	20	30	40	50	60
orf102ng	MMFSWFKLFHLFFVISWFAGLFYLPRI	10	20	30	40	50	60
orf102-1.pep	GFGAVVFGAAIPFAAGWWSGWVHV	70	80	90	100	110	120

-552-

```

      |||
orf102ng  GFGAVVFGAAIPFAAGRWGSGVHVHVKLCLGLMLLAYQLYCGVLLRRRFQDYSNAFSHRWYR
              70          80          90          100          110          120

              130          140
5  orf102-1.pep  VFNEIPVLLMVAALYLVVFKPFX
      |||
orf102ng  VFNEIPVLLMVAALYLVVFKPFX
              130          140

```

- 10 In addition, ORF102ng (SEQ ID NO: 764) shows significant homology to a membrane protein (SEQ ID NO: 1160) from *H.pylori*:

```

15 gi|2314656 (AE000647) conserved hypothetical integral membrane protein
   [Helicobacter pylori] Length = 148
   Score = 79.2 bits (192), Expect = 1e-14
   Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

Query: 3  FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
        F W K FH+ VISW A LFYLPR+ FV A + V++ +LY F++
Sbjct: 8  FLWVKAFHVI AVISWMAALFYLPRL FVYHAENAHKKEFVG VVQIQEK--KLYSFIAS PAM 65

20 Query: 63 GAVVFGAAIP-----FAAGRWGSGVHVHVKLCLGLMLLAYQLYCGVLLRRRFQDYSNAFS 115
        G + + F +G GW+H KL L ++LLAY YC +R + +
Sbjct: 66 GFTLITGILMLLIEPTL FKS G---GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121

Query: 116 HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
        R+YRVFNE P KPF
25 Sbjct: 122 ARFYRVFNEAPTILMILIVILVVVKPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 765>] (SEQ ID NO: 765):

```

30 1  ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CCGCGGCAGC
   51 GGT TTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
  101 TTACGGAAAC GGTCAGGCGC GGC // .....
  35 //... ATTCGTTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGGCGAAGCT
   51 CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGT TACAACA
  101 GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT
  151 GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
  201 GGT TGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
  251 TGAAAAATCG CCGCGGCAAG GCGTTTGTGC GCGTGTGGG TGCGGACGGC
  40 301 AAGGCGGCGG AACGCGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
   351 CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGGTC ATCTCCGAAA
   401 TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCCG
   451 CCGCGCCGAT AA

```


This corresponds to the amino acid sequence [<SEQ ID 766; ORF85>] (SEQ ID NO: 766; ORF85):

```

1  MAKMMKWA AV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....
5  51  .....
101 .....
151 .....
201 .....I SFTILSEPDT
251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARSFV PNPDGKLATG
301 MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM
10 351 RDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

```

Further work revealed the further partial nucleotide sequence [<SEQ ID 767>] (SEQ ID NO: 767):

```

1  ..GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
15 51  ACTCGGGCAA CAGGTTAAAA AGGGCGATTT GATTGCGGAA ATCAATTCTGA
101 CCTCGCAGAC CAATACGCTC AATACGAAA AATCCAAGTT GGAAACGTAT
151 CAGGCGAAGC TGGTGTCCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
201 ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAAGAGG
251 ATTTGGAAG CGCGCAGGAT GCGTTTGCCG CCGCCAAAGC CAATGTTGCC
301 GAGCTGAAGG CTTTAATCAG ACAGAGCAAA ATTTCCATCA ATACCGCCGA
20 351 GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
401 TGGCGATTCT CGTGAAGAG GGCAGACTG TGAACGCGGC GCAGTCTACG
451 CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAATGCA
501 GATTGCCGAG GCGGATATTA CCAAGGTGAA GGCGGGGCAG GATATTTCTG
551 TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC
25 601 GTCGACCCCG GGCTGACCAC GATGTCGTCG GGCGGTTACA ACAGCAGTAC
651 GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTCT TTTGTGCCGA
701 ATCCGGACGG CAAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
751 ATCGACGGCG TGAAAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA
30 801 TCGCGGCGGC AAGGCGTTTG TCGCGTGTT GGGTGCGGAC GGCAAGGCGG
851 CGGAACGCGA AATCCGGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA
901 AAAAGCGGGT TGAAAGAGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
951 CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CTTAGGCGGC CCGCCGCGCC
1001 GATAA

```

35 This corresponds to the amino acid sequence [<SEQ ID 768; ORF85-1>] (SEQ ID NO: 768; ORF85-1):

```

1  ..VSVGAQASGQ IKILYVKLGQ QVKKGDLIAE INSTSQNTL NTEKSKLETY
40 51  QAKLVSAQIA LGSAEKKYKR QAALWKENAT SKEDLESAQD AFAAAKANVA
101 ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST
151 PTIVQLANLD MMLNMQIAE GDITKVKAGQ DISFTILSEP DTPIKAKLDS
201 VDPGLTTMSS GGYNSSTDTA SNAVYYARS FVPNPDGKLA TGMTTQNTVE
251 IDGVKNVLI PSLTVKNRGG KAFVRLGAD GKAAEREIRT GMRDSMNTEV
301 KSGLKEGDKV VISEITAAEQ QESGERALGG PPRR*

```

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 (SEQ ID NO: 766) shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) (SEQ ID NO: 770) from strain A of *N. meningitidis*:

```

5      10      20      30      40
orf85.pep  MAKMMKWAAVA AAVAAA VGGWS-LKPEPHVLDITETVRRG
          |||||
orf85a     MAKMMKWAAVA AAVAAA VGGWSY LKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS
          10      20      30      40      50      60
                        //

10     80      90      100
orf85.pep  .....ISFTILSEPDTPIKAKLDSVDPGLTTMSSG
          |||||
orf85a     TIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSSG
          210     220     230     240     250     260

15     110     120     130     140     150     160
orf85.pep  GYNSSTD TASNAV VYYYAR SFVPNP DGKLAT GMTTQ NTVEID GVKNVL IIPSLT VKNRGG K
          |||||
orf85a     GYNSSTD TASNAV VYYYAR SFVPNP DGKLAT GMTTQ NTVEID GVKNVL IIPSLT VKNRGG R
          270     280     290     300     310     320

20     170     180     190     200     210     220
orf85.pep  AFVRVLG ADGKAA EIRTM RDSMN TEVKSG LKEGDK VVISEI TAAEQQ ESGERA LGGP
          |||||
orf85a     AFVRVLG ADGKAA EIRTM RDSMN TEVKSG LKEGDK VVISEI TAAEQQ ESGERA LGGP
          330     340     350     360     370     380

25     230
orf85.pep  PRRX
          |||
orf85a     PRRX
          390

```

30 The complete length ORF85a nucleotide sequence [<SEQ ID 769>] (SEQ ID NO: 769) is:

```

1  ATGGCAAAAA TGATGAAATG GCGCGCTGTT GCGGCGGTCTG CGGCGGCAGC
51 GGT TTGGGGC GGATGGTCTT ATCTGAAGCC CGAGCCGCAG GCTGCTTATA
101 TTACGGAAAC GGT CAGGCGC GCGGACATCA GCCGACGGT TTCTGCAACA
151 GCGGAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCATCGGG
35 201 GCAGATTAAG AAAC TTTATG TCAAAC TCGG GCAACAGGTT AAAAAGGGCG
251 ATTTGATTGC GGAAATCAAT TCGACCTCGC AGACCAATAC GCTCAATACG
301 GAAAAATCCA AATTGGAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
401 AGGATGATGC GACCGCTAAA GAAGATTGG AAAGCGCACA GGATGCGCTT
40 451 GCCGCCGCCA AAGCCAATGT TGCCGAGCTG AAGGCTCTAA TCAGACAGAG
501 CAAAATTTC ATCAATACCG CCGAGTCGGA ATTGGGCTAC ACGCGCATT
551 CCGCAACGAT GGACGGCAGC GTGGTGGCGA TTCTCGTGGA AGAGGGGCAG
601 ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
45 701 TGAAGGCGGG GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGATACG
751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTACT
851 ATTATGCCCG TCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
901 ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGCTGAT
50 951 TATTCGCTCG CTGACCGTGA AAAATCGCGG CGGCAGGGCG TTTGTGCGCG
1001 TGT TGGGTGC AGACGGCAAG GCGGCGGAAC GCGAAATCCG GACCGGTATG
1051 AGAGACAGTA TGAATACCGA AGTAAAAAGC GGGTTGAAAG AGGGGGACAA

```

1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
1151 GCGCCCTAGG CGGCCCGCCG CGCCGATAA

This encodes a protein having amino acid sequence [<SEQ ID 770>] (SEQ ID NO: 770):

5	1	MAKMMKWA	AAV	AAVAAA	AVWG	GWSYLKPEPQ	AA	YITETVRR	GDISRTVSAT
	51	GEISPSNLVS	VGAQASGQIK	KLYVKLGQQV	KKGDLIAEIN	STSQTNTLNT			
	101	EKSKLETYQA	KLVSAQIALG	SAEKKYKRQA	ALWKDDATAK	EDLESAQDAL			
	151	AAAKANVAEL	KALIRQSKIS	INTAESELGY	TRITATMDGT	VVAILVEEGQ			
	201	TVNAAQSTPT	IVQLANLDM	LNKMQIAEGD	ITKVKAGQDI	SFTILSEPDT			
10	251	PIKAKLDSVD	PGLTTMSSGG	YNSSTDASN	AVYYIARSFV	PNPDGKLATG			
	301	MTTQNTVEID	GVKNVLIIPS	LTVKNRGGRA	FVRVLGADGK	AAEREIRTGM			
	351	RDSMNTVEKS	GKKEGDKVVI	SEITAEQQE	SGERALGGPP	RR*			

ORF85a (SEQ ID NO: 770) and ORF85-1 (SEQ ID NO: 768) show 98.2% identity in 334 aa overlap:

		30	40	50	60	70	80	
	orf85a.pep	PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE						
20	orf85-1				VSVGAQASGQIKILYVKLGQQVKKGDLIAE			
					10	20	30	
		90	100	110	120	130	140	
	orf85a.pep	INSTSQTNLTNTEKSKLETYQAKLVSAQIALGSAEKKYKQQAALWKDDATAKEDLESAQD						
25	orf85-1	INSTSQTNLTNTEKSKLETYQAKLVSAQIALGSAEKKYKQQAALWKENATSKEDLESAQD						
			40	50	60	70	80	90
		150	160	170	180	190	200	
	orf85a.pep	ALAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST						
30	orf85-1	AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST						
			100	110	120	130	140	150
		210	220	230	240	250	260	
	orf85a.pep	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS						
35	orf85-1	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS						
			160	170	180	190	200	210
		270	280	290	300	310	320	
	orf85a.pep	GGYNSSTD TASNAVYY YARSFV PNPDGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG						
40	orf85-1	GGYNSSTD TASNAVYY YARSFV PNPDGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG						
			220	230	240	250	260	270
		330	340	350	360	370	380	
	orf85a.pep	RAFV RVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKV VISEITAAEQQESGERALGG						
		:						
45	orf85-1	KAFV RVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKV VISEITAAEQQESGERALGG						
			280	290	300	310	320	330
		390						
	orf85a.pep	PPRRX						
	orf85-1	PPRRX						

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a (SEQ ID NO: 770).

Homology with a predicted ORF from *N.gonorrhoeae*

- 5 ORF85 (SEQ ID NO: 766) shows a high degree of identity with a predicted ORF (ORF85ng) (SEQ ID NO: 772) from *N.gonorrhoeae*:

ORF85	1	MAKMMKWA AVAAVAAA VWGWS.LKPEPHVLDITETVRRG.....	40
ORF85ng	1	MAKMMKWA AVAAVAAA VWGWSY LKPEPQAAYITEAVRRGDISRTVSAT	50
ORF85	ISFTILSEPDT	250
ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDT	250
ORF85	251	PIKAKLDSVDPLTTMSSGGYNSSTD TASNAVYYAR SFVNPDPGKLATG	300
ORF85ng	251	PIKAKLDSVDPLTTMSSGGYNSSTD TASNAVYYAR SFVNPDPGKLATG	300
ORF85	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
ORF85ng	301	MTTQNTVEIDGVKNVLLIPSLTVKNRGGKAFVRVLGADGKAVEREIRTGM	350
ORF85	152	RDSMNTVEKSGLKEGDKV VISEITAAEQQESGERALGGPPRR	393
ORF85ng	351	KDSMNTVEKSGLKEGDKV VISEITAAEQQESGERALGGPPRR	393

The complete length ORF85ng nucleotide sequence [<SEQ ID 771>] (SEQ ID NO: 771) is:

1	ATGGCAAAA	TGATGAAATG	GGCGGCTGTT	GCGGCGGTCG	CGGCGGCaac
51	GGTTTGGGCG	GGATGGTCTT	ATCTGAAGCC	CGAACCGCAG	GCTGCTTATA
101	TTACGGAAac	ggTCAGGCGC	GGCGATATCA	GCCGACGGT	TTCCGCGACG
151	GgcgAGATT	CGCCGTCCAA	CCTGGTATCG	GTCGGCGCGC	AGGCTTCGGG
201	GCAGATTAAA	AAGCTTTATG	TCAAACCTCGG	GCAACAGGTC	AAAAAGGGCG
251	ATTTGATTGC	GGAAATCAAT	TCGACCACGC	AGACCAACAC	GATCGATATG
301	GAAAAATCCA	AATTGGAAAC	GTATCAGGCG	AAGCTGGTGT	CGGCACAGAT
351	TGCATTGGGC	AGCGCGGAGA	AGAAAATATA	GCGTCAGGCG	GCGTTGTGGA
401	AGGATGATGC	GACCTCTAAA	GAAGATTGG	AAAGCGCGCA	GGATGCGCTT
451	GCCGCCGCCA	AAGCCAATGT	TGCCGAGTTG	AAGGCTTTAA	TCAGACAGAG
501	CAAAATTTC	ATCAATACCG	CCGAGTCGGA	TTTGGGCTAC	ACGCGCATTA
551	CCGCGACGAT	GGACGGCAGC	GTGGTGCGCA	TTCCCGTGGA	AGAGGGGCAG
601	ACTGTGAACG	CGGCGCAGTC	TACGCCGACG	ATTGTCCAAT	TGGCGAATCT
651	GGATATGATG	TTGAACAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
701	TGAAGGCGGG	GCAGGATATT	TCGTTTACGA	TTTTGTCCGA	ACCGGATACG
751	CCGATTAAGG	CGAAGCTCGA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
801	CTCGGGCGGC	TACAACAGCA	GTACGGATAC	GGCTTCCAAT	GCGGTCTATT
851	ATTATGCCCG	TTTCGTTTGTG	CCGAATCCGG	ACGGCAAAC	CGCCACGGGG
901	ATGACGACGC	AGAATACGGT	TGAAATCGAC	GGTGTGAAAA	ATGTGTTGCT
951	TATTCCGTCG	CTGACCGTGA	AAAATCGCGG	CGGCAAGGCG	TTCGTACGCG
1001	TGTTGGGTGC	GGACGGCAAG	GCAGTGAAC	GCGAAATCCG	GACCGGTATG
1051	AAAGACAGTA	TGAATACCGA	AGTGAAGAGC	GGGTTGAAAG	AGGGGGACAA

1	MAKMMKWA	AAV	AAVAAA	AVWG	GWSY	LKPEPQ	AA	ITEAVRR	GD	ISRTVSAT
51	GEISPSNL	VS	VGAQ	ASGQIK	KLYV	KLGOQV	KK	GLIAEIN	ST	TQNTIDM
101	EKSKLETY	QA	KLVS	AQIALG	SAEK	KYKRQA	AL	WKDDATSK	ED	LESAQDAL
151	AAAKANVA	EAL	KALIR	QSKIS	INTA	ESDLGY	TR	ITATMDGT	VV	AIPVEEGQ
201	TVNAAQST	PPT	IVQL	ANLDM	LNM	QIAEGD	IT	KVKAGQDI	SF	TILSEPDT
251	PIKAKLDS	VD	PGLT	MTSSGG	YNS	STDASN	AV	YYARSFV	PN	PDGKLGATG
301	MTTQNTVE	ID	GVKN	VLLIPS	LTVK	NRGGKA	FV	RVLGADGK	AV	EIRITGM
351	KDSMNT	TEVKS	GLKE	GDKVVI	SE	ITAAEQQE	SG	ERALGGPP	RR	*

	30	40	50	60	70	80
orf85ng	PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE					
orf85-1	VSVGAQASGQIKILYVKLGQQVKKGDLIAE					
	10	20	30			
	90	100	110	120	130	140
orf85ng	INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD					
orf85-1	INSTSQTNTLNTTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD					
	40	50	60	70	80	90
	150	160	170	180	190	200
orf85ng	ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST					
orf85-1	AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST					
	100	110	120	130	140	150
	210	220	230	240	250	260
orf85ng	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS					
orf85-1	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS					
	160	170	180	190	200	210
	270	280	290	300	310	320
orf85ng	GGYNSSTD TASNAVYYYARSFV PNP DGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG					
orf85-1	GGYNSSTD TASNAVYYYARSFV PNP DGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG					
	220	230	240	250	260	270
	330	340	350	360	370	380
orf85ng	KAFV RVLGADGKAVEREIRTGMKDSMNTVEVKSGLKEGDKVVISEITAAEQQESGERALGG					
orf85-1	KAFV RVLGADGKAAEREIRTGM RDSMNTVEVKSGLKEGDKVVISEITAAEQQESGERALGG					
	280	290	300	310	320	330
	390					
orf85ng	PPRRX					
orf85-1	PPRRX					

In addition, ORF85ng (SEQ ID NO: 772) shows significant homology to an *E.coli* membrane fusion protein (SEQ ID NO: 1161):

```

5      gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from membrane
      fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia coli] Length
      = 380
      Score = 193 bits (485), Expect = 2e-48
      Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

10     Query: 29  PQAAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE 88
      P   Y T  VR GD+ ++V ATG++      V VGAQ SGQ+K L V +G +VKK  L+
      Sbjet: 41  PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTLSPAIGDKVKKDQLLG 100

      Query: 89  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKYKRQAALWKDDATSKEXXXXXXX 148
      I+   N I   ++ L  +A+   A+ L  A   Y RQ  L +   A S++
      Sbjet: 101 IDPEQAENQIKEVEATLMELRAQRQAEAEELKLARVTYSRQRLAQTKAVSQQLDLDTAAT 160

15     Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST 208
      I++++ S++TA+++L YTRI A M G V  I   +GQTV AAQ
      Sbjet: 161 EMAVKQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQA 220

      Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKGAGDISFTILSEPDTPIKAKLDSVDPLTMTSS 268
      P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++ V P
      Sbjet: 221 PNILTLADMSAMLVKAQVSEADVHLKPGQKAWFTVLGDPLTRYEGQIKDVLP----- 273

20     Query: 269 GGYNSSTDTASNAYYYYARSFVPNPDGKLTGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328
      + +  ++A++YYAR VPNP+G L  MT Q  +++ VKNVL IP  + + G
      Sbjet: 274 -----TPEKVNDAlFYARFEVPPNPNGLLRDLMTAQVHIQLTDVKNVLTIPSLALGDPVG 328

      Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVI 372
      +V L  +G+  ERE+  G ++  + E+  GL+  GD+VVI E
25     Sbjet: 329 DNRYKVKLLRNGETREREVTIGARNDDVEIVKGLEAGDEVVIGE 373

```

Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (SEQ ID NO: 768) (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 (SEQ ID NO: 768) is a surface-exposed protein, and that it is a useful immunogen.

35 Example 92

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 773] (SEQ ID NO: 773):

1 ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
 51 TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
 101 CGGTTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
 151 GGCAAACTGT ATGCGGAAGc CAAATTCGCC GACgGcAGCG TAACTTACGG
 5 201 CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAAGGCT ATGGATTTGT
 251 TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCGGGG
 301 CTGAAAATCA CCAACGGCAA AAAACTTTAT TCCGTCGGCG GTTTGAATAA
 351 GCGGGGTACA GAAAAATACA GCATAGGCGG CGTGGAACC GAAGTCGTCA
 401 AATATCGGGT GCGGCGCGGC GACGATGCGG TAATGTATTT cTTCGCACCG
 10 451 TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAAC
 501 CTATACGCTG AAACCTCAAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC
 551 CGTAA

This corresponds to the amino acid sequence [<SEQ ID 774; ORF120>] (SEQ ID NO: 774;
 15 ORF120):

1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR
 51 GKLYAEAKFA DGSVTYKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
 101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP
 20 151 SLNNIPAQIG YTDDGKTYTL KLKSVQINGQ AAKP*

Further work revealed the complete nucleotide sequence [<SEQ ID 775>] (SEQ ID NO: 775):

1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCGGC
 51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT
 101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
 25 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
 201 TTTTCGAGTCC GCGGTACGG TTGTTCGGCAA TACCCTGCAC CCTACCTACT
 251 ATAGAGACAT ACGCAGGGGC AAACGTGTAT CGGAAGCCAA ATTCGCCGAC
 301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
 351 CAAGGCTATG GATTTGTTCA CGCTTGCTTG GCAGTTGGCG GCAAATGACG
 30 401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
 451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
 501 GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
 551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
 601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
 35 651 CGGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence [<SEQ ID 776; ORF120-1>] (SEQ ID NO: 776;
ORF120-1):

1 MMKTFKNIFS AAILSAAALPC AYAAGLPQSA VLHYSYSGYI PATMTFERSG
 40 51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
 201 TDDGKTYTLK LKSVQINGQA AKP*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 (SEQ ID NO: 774) shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) (SEQ ID NO: 778) from strain A of *N. meningitidis*:

```

5      orf120.pep      10      20      30      40      50      60
      IPATMTFERSGNAYKIVSTIKVPLYNIRFE
      |||| : || |||||
orf120a      SAAILSAAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIKVPLYNIRFE
      10      20      30      40      50      60

10     orf120.pep      40      50      60      70      80      90
      SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAGESKTEQSPKAMD LFTLAWQL
      ||||| : |||||
orf120a      SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAXXXXXXQSPKAMD LFTLAWQL
      70      80      90      100     110     120

15     orf120.pep      100     110     120     130     140     150
      AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP
      ||||| : |||||
orf120a      AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP
      130     140     150     160     170     180

20     orf120.pep      160     170     180
      SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX
      ||||| : |||||
orf120a      SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX
      190     200     210     220

```

The complete length ORF120a nucleotide sequence [<SEQ ID 777>] (SEQ ID NO: 777) is:

```

25      1  ATGATGAAGA  CTTTTAAAAA  TATATTTTCC  GCCGCCATTT  TGTCCGCCGC
      51  CCTGCCGTGC  GCGTATGCGG  CAGGGCTGCC  CNAATCCGCC  GTGCTGCACT
      101 ATTCCGGCAG  CTACGGCATT  CCCGCCACNA  NNANNTNNGN  ACNNNGNGNC
      151 AATGCTTNCA  AAATCGTTTC  GACGATTAAA  GTGCCGCTAT  ACAATATCCG
      201 TTTCGAGTCC  GCGGTACGG  TTGTCGGCAA  TACCCTGCAC  CCTACCTACT
30     251 ATAGAGACAT  ACGCAGGGGC  AAAGTGTATG  CGGAAGCCAA  ATTCGCCGAC
      301 GGCAGCGTAA  CCTACGCCAA  AGCGGNNNNN  ANCNNNNNG  NGCAAAGCCC
      351 CAAGGCTATG  GATTTGTTCA  CGCTTGCNTG  GCAGTTGGCG  GCAAATGACG
      401 CGAAACTCCC  CCCGGGGCTG  AAAATCACCA  ACGGCAAAAA  ACTTTATTCC
      451 GTCGGCGGTT  TGAATAAGGC  GGGTACAGGA  AAATACAGCA  TAGGCGGCGT
35     501 GGAAACCGAA  GTCGTCAAAT  ATCGGTGCG  GCGCGGCGAC  GATGCGGTAA
      551 TGTATTTCTT  CGCACCGTCC  CTGAACAATA  TTCCGGCACA  AATCGGCTAT
      601 ACCGACGACG  GCAAACCTA  TACGCTGAAA  CTCAAATCGG  TGCAGATCAA
      651 CGGCCAGGCA  GCCAAACCGT  AA

```

40 This encodes a protein having amino acid sequence [<SEQ ID 778>] (SEQ ID NO: 778):

```

45      1  MMKTFKNIFS  AAILSAAALPC  AYAAGLPXSA  VLHYSGSYGI  PATXXXXXXX
      51  NAXKIVSTIK  VPLYNIRFES  GGTVVGNTLH  PTYYRDIRRG  KLYAEAKFAD
      101 GSVTYGKAXX  XXXXQSPKAM  DLFTLAWQLA  ANDAKLPPGL  KITNGKKLYS
      151 VGGLNKAGTG  KYSIGGVETE  VVKYRVRRGD  DAVMYFFAPS  LNNIPAQIGY
      201 TDDGKTYTLK  LKSVQINGQA  AKP*

```

ORF120a (SEQ ID NO: 778) and ORF120-1 (SEQ ID NO: 776) show 93.3% identity in 223 aa overlap:

-561-

		10	20	30	40	50	60
	orf120a.pep	MMKTFKNIFSAAILSAALPCAYAAGLPXS	AVLHYS	SGSYGIPAT	XXXXXXXX	NAXKIVSTIK	
5	orf120-1	MMKTFKNIFSAAILSAALPCAYAAGLPQSA	VLHYS	SGSYGIPAT	MTFERSGN	AYKIVSTIK	
		10	20	30	40	50	60
	orf120a.pep	VPLYNIRFESGGTVVGNTLHPTTYRDIRRG	KLYAEAKFADGS	VTY	GKAXXXXXX	QSPKAM	
10	orf120-1	VPLYNIRFESGGTVVGNTLHPTTYRDIRRG	KLYAEAKFADGS	VTY	GKAGESKTE	QSPKAM	
		70	80	90	100	110	120
	orf120a.pep	DLFTLAWQLAANDAKLPPGLKITNGKKLYS	VGGLNKAGT	GKYSIGGV	ETEVVKYR	VRRGD	
15	orf120-1	DLFTLAWQLAANDAKLPPGLKITNGKKLYS	VGGLNKAGT	GKYSIGGV	ETEVVKYR	VRRGD	
		130	140	150	160	170	180
	orf120a.pep	DAVMYFFAPSLNNIPAQIGYTDDGKTYTL	KLKSVQING	QAAK	PKX		
20	orf120-1	DAVMYFFAPSLNNIPAQIGYTDDGKTYTL	KLKSVQING	QAAK	PKX		
		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 (SEQ ID NO: 774) shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) (SEQ ID NO: 780) from *N.gonorrhoeae*:

25	orf120.pep	IPATMTFERSGNAYKIVSTIKVPLYNIRFE	30
	orf120ng	SAAILSAAALPCAYAARLPQSAVLHYS	69
	orf120.pep	SGGTVVGNTLHPTTYRDIRRGKLYAEAKFADGS	90
	orf120ng	SGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGS	129
30	orf120.pep	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGV	150
	orf120ng	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGV	189
	orf120.pep	SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP	184
35	orf120ng	SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP	223

The complete length ORF120ng nucleotide sequence [<SEQ ID 779>] (SEQ ID NO: 779) is:

	1	ATGATGAAGA	CTTTTAAAAA	TATATTTTCC	GCCGCCATTT	TGTCCGCCGC	
40	51	CCTGCCGTGC	GCGTATGCGG	CAAGGCTACC	CCAATCCGCC	GTGCTGCACT	
	101	ATTCCGGCAG	CTACGGCATT	CCCGCCACGA	TGACATTTGA	ACGCAGCGGC	
	151	AATGCTTACA	AAATCGTTTC	GACGATTAAA	GTGCCGCTAT	ACAATATCCG	
	201	TTTCGAATCC	GGCGGTACGG	TTGTCGGCAA	TACCCTGCAC	CCTGCCTACT	
	251	ATAAAGACAT	ACGCAGGGGC	AAACTGTATG	CGGAAGCCAA	ATTCCGCCGAC	
	301	GGCAGCGTAA	CCTACGGCAA	AGCGGGCGAG	AGCAAACCG	AGCAAAGCCC	
45	351	CAAGGCTATG	GATTTGTTCA	CGCTTGCCCTG	GCAGTTGGCG	GCAATGACG	

5
401 CGAAACTCCC CCCGGGTCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
451 GTCGGCGGCC TGAATAAGGC GGTACGGGA AAATACAGCA TaggCGGCGT
501 GGAAACCGAA GTCGTCAAAT ATCGGGTGC GCGCGGCGAC GATACGGTAA
551 CGTATTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601 ACCGACGACG GCAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA
651 CGGACAGGCC GCCAAACCGT AA

This encodes a protein having amino acid sequence [SEQ ID 780] (SEQ ID NO: 780):

10
1 MMKTFKNIFS AAILSAAALPC AYAARLPQSA VLHYSGSYGI PATMTFERSG
51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PAYYKDIRRG KLYAEAKFAD
101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

15 In comparison with ORF120-1 (SEQ ID NO: 776), ORF120ng (SEQ ID NO: 780) shows 97.8% identity in 223 aa overlap:

20
orf120-1.pep 10 20 30 40 50 60
MMKTFKNIFS AAILSAAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
orf120ng 10 20 30 40 50 60
MMKTFKNIFS AAILSAAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
25
orf120-1.pep 70 80 90 100 110 120
VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
orf120ng 70 80 90 100 110 120
VPLYNIRFESGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
30
orf120-1.pep 130 140 150 160 170 180
DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
orf120ng 130 140 150 160 170 180
DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
35
orf120-1.pep 190 200 210 220
DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX
orf120ng 190 200 210 220
DTVTYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX

40 This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* [SEQ ID 781] (SEQ ID NO: 781):

-563-

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
 51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
 101 CTCCGTTTGC GGTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
 5 201 GATGGTGTTC TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC
 251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTCAGGCGC
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
 10 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence [<SEQ ID 782; ORF121>] (SEQ ID NO: 782; ORF121):

1 MYRRKGRGIK PWMGAGXafa ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
 15 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
 151 RQGGNI..

Further work revealed the complete nucleotide sequence [<SEQ ID 783>] (SEQ ID NO: 783):

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
 20 51 GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
 101 CTCCGTTTGC GGTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
 201 GATGGTGTTC TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC
 25 251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTCAGGCGC
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC
 501 CTTGCTGCTT TACTATTTC TGTGATTG GCAGCGGTGG TCGTGGGCA
 30 551 TTGCCAAACT GGTTCGAAGG CGTTTGGCCG GTGCTTATAC GCGCATTACA
 601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT
 651 AATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGGTG CTGGTCGGGC
 701 TGGATTCGGG GTTTGCCATC GGTATGCTTG CCGTATTTT GGTGTTTGT
 751 CCTTATCTCG GGGCGTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCCTT
 35 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG
 851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACCCGAA AATCGTGGGA
 901 GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT
 951 CGGGCAGCTG ATGGGCTTGT TCGGAATGTT GCGGGGATTG CCTTTGGCCG
 1001 CCGTAACCTT GGTCTTGCTT CCGAGGGCG TGCAGAAATA TTTTGGCCGC
 40 1051 AGTTTTTACC GGGCAGGTA G

This corresponds to the amino acid sequence [<SEQ ID 784; ORF121-1>] (SEQ ID NO: 784; ORF121-1):

1 MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
 45 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
 151 RQGGNIVSSI GNLLLLPLLL YYFLDWQRW SCGIKLVPR RFAGAYTRIT
 201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMLAGILVFE
 251 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG
 301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAATLVLL REGVQKYFAG
 50 351 SFYRGR*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF121 (SEQ ID NO: 782) shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) (SEQ ID NO: 786) from strain A of *N. meningitidis*:

5	orf121.pep	10	20	30	40	50	60
		MYRRKGRGIKPWMGAGXAF	AAALVVLV	FALGDTLTPFAVA	AVLAYVLDPL	VEWLQKKGLNR	
	orf121a	MYRRKGRGIKPWMDAGAA	FAALVVLV	FALGDTLTPFAVA	AVLAYVLDPL	VEWLQKKGLNR	
10	orf121.pep	70	80	90	100	110	120
		ASASMSVMVFS	LILLALLLI	IVPMLVGQFNN	LASRLPQLIGFM	QNTLLPWLKNT	IGGYV
	orf121a	ASASMSVMVFS	LILLALLLI	IVPMLVGQFNN	LASRLPQLIGFM	QNTLLPWLKNT	IGGYV
15	orf121.pep	130	140	150			
		EIDQASIIA	WLQAHTGELS	NALKAWFPVLM	RQGGNI		
	orf121a	EIDQASIIA	WLQAHTGELS	NALKAWFPVLM	RQGGNIVSSIGN	LLLLPLLLYY	FLLDWQRW
20	orf121a	190	200	210	220	230	240
		SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGE	FLRGQLLVML	IMGLVYGLGL	VLVGLD

The complete length ORF121a nucleotide sequence [<SEQ ID 785>] (SEQ ID NO: 785) is:

25	1	ATGTATCGGA	GGAAAGGGCG	GGGCATCAAG	CCGTGGATGG	ATGCCGGTGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTATTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGTCTGT
	201	GATGGTGTTT	TCCTTGATTT	TGTTGTTGGC	ATTATTGTTG	ATTATTGTCC
30	251	CTATGCTGGT	CGGGCAGTTC	AACAATTTGG	CATCGCGCCT	GCCCCAATTA
	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	CTTCAGGCGC
	401	ATACGGGCGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
	451	AGGCAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCTTCC
35	501	CTTGCTGCTT	TACTATTTCC	TGCTGGATTG	GCAGCGGTGG	TCGTGCGGCA
	551	TTGCCAAACT	GGTTCCGAGG	CGTTTTGCCG	GTGCTTATAC	GCGCATTACA
	601	GGCAATTTGA	ACGAGGTATT	GGGCGAATTT	TTGCGCGGGC	AGCTTCTGGT
	651	GATGCTGATT	ATGGGTTTGG	TTTACGGCTT	GGGGTTGGTG	CTGGTCGGGC
	701	TGGATTCGGG	GTTTGCAATC	GGTATGGTTG	CCGGTATTTT	GTTTTTGTG
	751	CCCTATTGGG	GCGCGTTTAC	AGGACTGCTG	CTGGCAACCG	TCGCCGCCTT
40	801	GCTCCAGTTC	GGTTCGTGGA	ACGGCATCTT	GGCTGTTTGG	GCGGTTTTTG
	851	CCGTAGGACA	GTTTCTCGAA	AGTTTTTTCA	TTACGCCGAA	AATCGTGGGA
	901	GACCGTATCG	GCCTGTGCGC	GTTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
	951	CGGGCAGCTG	ATGGGCTTTG	TCGGAATGTT	GGCCGGATTG	CCTTTGGCCG
45	1001	CCGTAACCTT	GGTCTTGCTT	CGCGAGGGCG	TGCAGAAATA	TTTTGCCGGC
	1051	AGTTTTTACC	GGGGCAGGTA	G		

This encodes a protein having amino acid sequence [<SEQ ID 786>] (SEQ ID NO: 786):

1 MYRRKGRGIK PWMDAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM
 151 RQGGNIVSSI GNLLLLPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
 201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMVAGILVVF
 251 PYLGAFITGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
 301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG
 351 SFYRGR*

10 ORF121a (SEQ ID NO: 786) and ORF121-1 (SEQ ID NO: 784) show 99.2% identity in 356 aa overlap:

		10	20	30	40	50	60
	orf121a.pep	MYRRKGRGIK PWMDAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV EWLQKKGLNR					
15	orf121-1	MYRRKGRGIK P W M G A G A A F A A L V W L V F A L G D T L T P F A V A A V L A Y V L D P L V E W L Q K K G L N R					
		10	20	30	40	50	60
	orf121a.pep	ASASMSVMVFS L I L L L A L L L I I V P M L V G Q F N N L A S R L P Q L I G F M Q N T L L P W L K N T I G G Y V					
20	orf121-1	ASASMSVMVFS L I L L L A L L L I I V P M L V G Q F N N L A S R L P Q L I G F M Q N T L L P W L K N T I G G Y V					
		70	80	90	100	110	120
	orf121a.pep	EIDQASIIAW LQAHTGELSN ALKAWFPVLM RQGGNIVSS I G N L L L L P L L L Y Y F L L D W Q R W					
25	orf121-1	EIDQASIIAW LQAHTGELSN ALKAWFPVLM RQGGNIVSS I G N L L L L P L L L Y Y F L L D W Q R W					
		130	140	150	160	170	180
	orf121a.pep	SCGIAKLVPR RFAGAYTRIT GNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI					
30	orf121-1	SCGIAKLVPR RFAGAYTRIT GNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI					
		190	200	210	220	230	240
	orf121a.pep	GMVAGILVFPY LGAFITGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG					
35	orf121-1	GMLAGILVFPY LGAFITGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG					
		250	260	270	280	290	300
	orf121a.pep	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX					
40	orf121-1	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX					
		310	320	330	340	350	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 (SEQ ID NO: 782) shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) (SEQ ID NO: 788) from *N.gonorrhoeae*:

```

orf121.pep  MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR  60
             |||||:|||||
orf121ng    MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAYVLDPLVEWLQKKGLNR  60

orf121.pep  ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV  120
             |||||:|||||
orf121ng    ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV  120

orf121.pep  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI  156
             |||||:|||||
orf121ng    EIDQASIIAWFQAHTGELSNAKAWFPVLMKQGGNIVSTIGNLLLPPLLYYFLLDWHRW  180

```

An ORF121ng nucleotide sequence [<SEQ ID 787>] (SEQ ID NO: 787) was predicted to encode a protein having amino acid sequence [<SEQ ID 788>] (SEQ ID NO: 788):

```

1  MYRRKGRGIK PWMGAGAAFA ALVWLVYALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLLALLL IIIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
151 KQGGNIVSTI GNLLPPLLL YYFLLDWHRW SCGIPKLVPR RFAGAYTRIT
201 GNLNKVWGKF LRGQLLGETE RGAUVCRVGR ECWEGGGARS RPSDDGWPRW
251 GGG*

```

Further work revealed the following gonococcal DNA sequence [<SEQ ID 789>] (SEQ ID NO: 789):

```

1  ATGTATCGGA GAAAAGGACG GGGCATCAAG CCGTGGATGG GTGCCGGCGC
51 GCGTTTTGCC GCCTTGGTCT GGCTGGTTTA CGCGCTCGGC GATACTTTGA
101 CTCCGTTTGC GGTTCGCGCG GTGCTGGCGT ATGTGTTGGA CCCTTTGGTC
151 GAATGGTTGC AGAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
201 GATGGTGTTT TCCTTGATT TGTGTGTGGC ATTATTGTTG ATTATTGTCC
251 CTATGCTGGT CGGGCAGTTC AATAATTGG CATCTCGCCT GCCCAATTA
301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG TTTCAGGCGC
401 ATACGGGCGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTGTATG
451 AAACAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCCGCC
501 CTTGCTGCTT TACTATTTC TGTGCGATTG GCAGCGGTGG TCGTGGCGCA
551 TCGCCAACT GGTTCGAGG CGTTTTCGCG GTGCTTATAC GCGCATACG
601 GGTAATTTGA ACGAGTATT GGGCGAATTT TTGCGCGGTC AGCTTCTGGT
651 GATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGATG CTAGTCGGAC
701 TGGATTCGGG ATTTGCCATC GGTATGGTTG CCGGTATTTT GGTGTTTGTC
751 CCCTATTTGG GTGCGTTTAC GGGATTGCTG CTTGCCACTG TTGCAGCCTT
801 GCTCCAGTTC GGTTCGTGGA ACGGAATCTT GGCTGTTTGG GCGGTTTTTG
851 CCGTCGGTCA GTTCTCGAA AGTTTTTTCA TTACGCCGAA AATTGTAGGA
901 GACCGTATCG GCCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT
951 CGGAGAGCTG ATGGGCTTTG TCGGAATGTT GGCCGGATTG CCTTTGCCCG
1001 CCGTAACCTT GGTCTTGCTT CGCGAGGGCG CGCAGAAATA TTTTGCCGGC
1051 AGTTTTTACC GGGGCAGGTA G

```

This corresponds to the amino acid sequence [<SEQ ID 790; ORF121ng-1>] (SEQ ID NO: 790; ORF121ng-1):

```

1  MYRRKGRGIK PWMGAGAAFA ALVWLVYALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLLALLL IIIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM

```

-567-

151 KQGGNIVSSI GNLLLPPLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
 201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLM LVGLDSGFAI GMVAGILVFV
 251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
 301 DRIGLSPFWV IFSLMAFGEL MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
 351 SFYRGR*

ORF121ng-1 (SEQ ID NO: 790) and ORF121-1 (SEQ ID NO: 784) show 97.5% identity in 356 aa overlap:

```

10      orf121-1.pep      10      20      30      40      50      60
      MYRRKGRGIKPWMGAGAAFAALVWLVFALGDTLTPFAVAAYLAYVLDPLVEWLQKKGLNR
      orf121ng-1      MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAYLAYVLDPLVEWLQKKGLNR
                        10      20      30      40      50      60

15      orf121-1.pep      70      80      90      100     110     120
      ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      orf121ng-1      ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
                        70      80      90      100     110     120

20      orf121-1.pep      130     140     150     160     170     180
      EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLLYYFLLDWQRW
      orf121ng-1      EIDQASIIAWFQAHTGELSNAKAWFPVLMKQGGNIVSSIGNLLLLPLLLLYYFLLDWQRW
                        130     140     150     160     170     180

25      orf121-1.pep      190     200     210     220     230     240
      SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI
      orf121ng-1      SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI
                        190     200     210     220     230     240

30      orf121-1.pep      250     260     270     280     290     300
      GMLAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
      orf121ng-1      GMVAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG
                        250     260     270     280     290     300

35      orf121-1.pep      310     320     330     340     350
      DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
      orf121ng-1      DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX
                        310     320     330     340     350

```

40 In addition, ORF121ng-1 (SEQ ID NO: 790) shows homology to a permease (SEQ ID NO: 1162) from *H. influenzae*:

sp|P43969|PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
 Score = 69.9 bits (168), Expect = 2e-11
 Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)

45 Query: 26 VYALGDTLTPFAVAAYLAYVLDPLVEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXXXV 84
 +Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
 Sbjct: 32 IYFFGDLIAPLLIALVLSYLLIIPINFLNQYLKCPRLATILIFGSFIGLAAVFFLVLP 91

Query: 85 MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNAK 143
 ML Q +L S LP + N WL N Y E ID + + + F + ++ +
 Sbjct: 92 MLWNQTISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFGE 147
 Query: 144 AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXWQWSCGIKLVPRRFAGAYTRITGNL 203

5 + + + N+VS D G+++ +P+ A+ R +
 Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRLPKNRNLAFXRWK-EM 206

Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXSGFAIGMVAGILVFVPYXXXXXXXXXXXX 263
 + + ++ G+ + + G+ V VPY
 Sbjct: 207 QQQISNYIHGKLEILIVTLITYIIFLIFGLNYPLLLAFVGLSVLVPIYIGAVIVTIPVA 266

10 Query: 264 XXXXXQFGSWNGILAVWAVFVAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQQLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326

Query: 324 VGMLAGLPLAAVTLVLL 340
 G+ +PLA + ++
 15 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 791>] (SEQ ID NO: 791):

25 1 ..ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
 51 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
 101 TTTGACGTC CTGCCGCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
 151 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
 201 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
 251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTCTGTGC
 30 301 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
 351 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
 401 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
 451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
 35 501 CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAG..

This corresponds to the amino acid sequence [<SEQ ID 792; ORF122>] (SEQ ID NO: 792; ORF122):

40 1 ..TAFSAALRLS PSXLVIFLSF GKPYYQTAAI LTFFCTSCPP RSNAQQYRR
 51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRRECGFLC
 101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
 151 EQRVGNGVQQ RIGIGVSEQP FFKWDFNSAK YQ..

Further work revealed the complete nucleotide sequence [<SEQ ID 793>] (SEQ ID NO: 793):

```

      1  ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGGGAAG TAGATACCGC
     51  GCCTTTGATT TTTTGGCCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA
    101  TGGTCGAGCC GGTACCGATG CCGATATATT CATTTTCGGG TACGAATTCG
5      151  ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTGTG TCGTCATATT
    201  TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
    251  TTTGCACGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
    301  CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
    351  TTTTGCCTTT GATGTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
10     401  ATGTTGGCAC GCATTGCGG AATGTGCGGC GCGAGTTTGG GTTCTGTGTC
    451  AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
    501  TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTGAACTCT
    551  GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
    601  GAGCAGCGCG TCGGTAACGG CGTGACAGCAG CGCATCGGCA TCGGAGTGTC
15     651  CGAGCAGCCC TTTTCAAAT GGGATTCAA CTCCGCCAAG TATCAGCTTT
    701  CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT
    751  CGTCATCGTT TGTGTTCTCT A

```

This corresponds to the amino acid sequence [<SEQ ID 794; ORF122-1>] (SEQ ID NO: 794:

20 ORF122-1):

```

      1  ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS
     51  TAFSAAMRLS SSCVVIFLSF GKPYQQTAAI LTFFCTSCPP RSNAYQQYRR
    101  LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRREFGFLC
    151  NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
25     201  EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDTDV
    251  RHRLCS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF122 (SEQ ID NO: 792) shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) (SEQ ID NO: 796) from strain A of *N. meningitidis*:

```

                                     10      20      30
or122.pep                          TAFSAALRLSPSXLVIFLSFGKPYQQTAAI
35  orf122a                          FLPLLPKASMKKLMEVPVPMPIYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI
                                     30      40      50      60      70      80

                                     40      50      60      70      80      90
or122.pep                          LTFFCTSCPPRSNAYQQYRRRLRYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR
40  orf122a                          LTFFXTSCPPRSNPYQQYRRRLRYAFHAPEITEFFVGFAFXVDARNVYAQIGGDVGTHLR
                                     90     100     110     120     130     140

                                     100     110     120     130     140     150
or122.pep                          NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT
45  orf122a                          NMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT
                                     150     160     170     180     190     200

```

orf122.pep 160 170 180
 EQRVGNQVQQRIGIGVSEQPFFKWFDFNSAKYQ
 orf122a EQRVGNQVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLVDIVALSDTDVRHRLCSX
 210 220 230 240 250

The complete length ORF122a nucleotide sequence [SEQ ID 795] (SEQ ID NO: 795) is:

1 ATATCATATT GGGCAAGCAG TTTACTGGAT TTTTGGGAAG TAGATACCGC
 51 GCCTTTGATT TTTTGGCCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA
 101 TGGTCGAACC GGTACCGATG CCGATGTATT CGTTTTCGGG TACGAATTCTG
 151 ACTGCNTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTGTT TCGTCATATT
 201 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
 251 TTNNNACGTC CTGCCCGCCG CGTTCAAATC CTTACCAGCA ATACCGCCGC
 301 CTGCGACTCT ATGCCTTCCA TGCGCCCGAG ATAACCGAGT TTTTCGTTGG
 351 TTTTGCCTTT GANGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
 401 ATGTTGGCAC GCATTGCGG AATATGCGGC GCGAGTTTGG GTTCTGTGTC
 451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
 501 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
 551 GCGGCGGTGT CGGGGAATG GCTGCGGATA TCGCCCAAAC CTGCCGCACC
 601 GAGCAGCGCG TCGGTAACGG CGTGACGAG CGCATCGGCA TCGGAGTGTC
 651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAT CTCCGCCAAG TATCAGCTTT
 701 CTGCCCTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT
 751 CGTCATCGTT TGTGTTCCCTG A

This encodes a protein having amino acid sequence [SEQ ID 796] (SEQ ID NO: 796):

1 ISYWASSSLD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PMYSFSGTNS
 51 TAFSAAMRLS SSCVVIPLSF GKPYQQTAAI LTFFXTSCPP RSNPYQQYRR
 101 LRLYAFHAP ETEFFVGFAF XVDARNVYQA IGGDVGTHLR NMRREFGFLC
 151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
 201 EQRVGNQVQQ RIGIGVSEQP FFKWFDFNSA YQLSAFGQLV DIVALSDTDV
 251 RHRLCS*

ORF122a (SEQ ID NO: 796) and ORF122-1 (SEQ ID NO: 794) show 96.9% identity in 256 aa overlap:

35 10 20 30 40 50 60
 orf122a.pep ISYWASSSLDFLEVDTAPLI FLPLLPKASM KKLMEVPVPM PMYSFSGTNS TAFSAAMRLS
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 orf122-1 ISYWASSSPDFLEVDTAPLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS TAFSAAMRLS
 10 20 30 40 50 60
 40 70 80 90 100 110 120
 orf122a.pep SSCVVIPLSFGKPYQQTAAI LTFFXTSCPP RSNPYQQYRR LRLYAFHAP ETEFFVGFAF
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 orf122-1 SSCVVIPLSFGKPYQQTAAI LTFFCTSCPP RSNAYQQYRR LRLYAFHPPEIAE FFFVGFAF
 70 80 90 100 110 120
 45 130 140 150 160 170 180
 orf122a.pep XVDARNVYQA IGGDVGTHLR NMRREFGFLC NHGRIDIDRL PTLRLNALIR RTQKDAAVRI
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 orf122-1 DVDARNVYQA IGGDVGTHLR NVREFGFLC NHGRIDIDRL PTLRLNALIR RTQKDAAVRI
 130 140 150 160 170 180
 50 190 200 210 220 230 240

-571-

orf122a.pep	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLV
orf122-1	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLV
	190 200 210 220 230 240

5

	250
orf122a.pep	DIVALSDDVRHRLCSX
orf122-1	DIVALSDDVRHRLCSX
	250

10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 (SEQ ID NO: 792) shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) (SEQ ID NO: 798) from *N.gonorrhoeae*:

	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
15	orf122ng	FLPLLPKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
	orf122ng	LTFFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
20	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
	orf122ng	NVRCEFGFLCNHGRIDIDHLPRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
	orf122.pep	EQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQ	182
25	orf122ng	EQRVGNVQQRVGIRMPEQPFFKWDFNSAKYQLSAFGQLVDIVALSDDIRHRLCS	256

The complete length ORF122ng nucleotide sequence [<SEQ ID 797>] (SEQ ID NO: 797) is:

	1	ATGTCGTACC	GGGCAAGCAG	TTCCGCCGAT	TTTTTGGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTTGCCCAA	GGCTTCGATG	AAGAAATTGa
30	101	tgGTCGAACC	GgtaCCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCTG
	151	ACTGCTTTT	CGGCGGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
	201	TTTAtccttt	gGGAAaccct	atcaAcaAAc	agccgccatC	TTAACATTTT
	251	TTTGCACGtc	ctggccgccg	cgttcaAATc	cgtaccaGca	ataccgccgc
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GATatTGACG	CACGAAATAT	CGatacCCAa	atcggcgGCG
35	401	ATGTTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACCTCT
	551	GCGGCGGTGT	CGGGAATG	GCTGCCGATG	TCGCCCCAAC	CTGCCGACCC
	601	GAGCAGCgcg	tcggtaaCGG	CGTGCAGCAG	cgcgTcgGCA	TCCGAATGCC
40	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
	751	CGTCATCGTT	TGTGTTCTTG	A		

This encodes a protein having amino acid sequence [<SEQ ID 798>] (SEQ ID NO: 798):

45

1 MSYRASSSPD FLEVETAPLI FLPLLPKASM KKLMEVPVPM PMYSFSGTNS

51 TAFSAAMRLS SSCVVIFLSF GKPYQQTAAI LTFFCTSWPP RSNPYQQYRR
 101 LRLYAFHPPE IAEFFVGFAF DIDARNIDTQ IGGDVGTHLR NVRCEFGFLC
 151 NHGRIDIDL PTLRLNALIR RTQKDAAVRI FELCGGVGKM AADVAQTCRT
 201 EQRVGNVQO RVGIRMPEQP FFKWDFNSAK YQLSAFGQLV DIVALSDDI
 251 RHRLCS*

ORF122ng (SEQ ID NO: 798) and ORF122-1 (SEQ ID NO: 794) show 92.6% identity in 256 aa overlap:

10	orf122-1.pep	10 20 30 40 50 60	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS
	orf122ng	10 20 30 40 50 60	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS
15	orf122-1.pep	70 80 90 100 110 120	SSCVVIFLSFGKPYQQTAAILTFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF
	orf122ng	70 80 90 100 110 120	SSCVVIFLSFGKPYQQTAAILTFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAF
20	orf122-1.pep	130 140 150 160 170 180	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI
	orf122ng	130 140 150 160 170 180	DIDARNIDTQIGGDVGTHLRNVRCEFGFLCNHGRIDIDHPLTLRLNALIRRTQKDAAVRI
25	orf122-1.pep	190 200 210 220 230 240	FELCGGVGEMAAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLV
	orf122ng	190 200 210 220 230 240	FELCGGVGKMAADVAQTCRTEQRVGNVQQRVGIRMPEQPFFKWDFNSAKYQLSAFGQLV
30	orf122-1.pep	250	DIVALSDDVHRRLCSX
	orf122ng	250	DIVALSDDIRHRLCSX

35 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 799>] (SEQ ID NO: 799):

40 1 ..GCCGGCGCGA GTGCGAACAA CATTTCCGCG CGTTTTCGCG AACACCCGT
 51 CGCTGTCAGC GTTACCCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCCG
 101 TTACCGAATA TGAAACTTC CTGCTGCTTA TCGGCTCGGT ATTTGCGCCG
 151 ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTTGAAAC GGCGTGA

This corresponds to the amino acid sequence [<SEQ ID 800; ORF125>] (SEQ ID NO: 800; ORF125):

```

1  ..AGASANNISA RFAETPVAVS VTLIGTVLAV MLPVTEYENF LLLIGSVFAP
51  MGGFDCRLFR LETA*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 801>] (SEQ ID NO: 801):

```

1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCTCCGCCA TCGGGCTGAT
51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
101  TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CGGCTCTACT TTTGGGTCAT
151  GCCGTCGGCG GCGCGCTGTT TTTTGCGGCG GCGTATATCG GCGCACTGAC
201  CGGACGCAGC TCGATGGAAA GCGTGCCTCT GTCGTTCCGGC AAACGCGGTT
251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
301  GTGATGATTT ACGCCGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
351  GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GGCGCGCTGA
401  TTGTGCTGTG GCTGGTTTTT GCGGCACGCA AAACAGGCGG GCTGAAAACC
451  GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAAGT
501  CTTTTCCACG GCAGGCAGCA CCGCCGCACA GGTTTCAGAC GGCATGAGTT
551  TCGGAACGGC AGTCGAGCTG TCCGCCGTGA TGCCGCTTTC CTGGCTGCCG
601  CTTGCCGCCG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCCT
651  GACGGCAACG CTCGCCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG
701  GTTTGGCAGC GCGGTTGTTT ACCGGAGAAA CCGACGTGGC AAAAATCCTG
751  CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTCT TCCTCTCCAC
801  CGTTACCACA ACGTTTCTCG ATGCCTATTC CGCCGCGCG AGTGCGAACA
851  ACATTTCCGC GCGTTTTCG GAAACACCCG TCGCTGTCGG CGTTACCCTG
901  ATCGGCACGG TACTTGCCGT CATGCTGCCC GTTACCGAAT ATGAAAACCT
951  CCTGCTGCTT ATCGGCTCGG TATTGCGCC GATGGCGGCG GTTTTGATTG
1001 CCGACTTTT CGTCTTGAAA CGGCGTGAGG AGATTGAAGG CTTTGACTTT
1051 GCCGGAAGT TTTGTTGGCT TCGGGGCTTC ATCCTCTACC GCTTCTGCT
1101 CTCGTCCGGC TGGGAAAGCA GCATCGGTCT GACCGCCCCC GTAATGTCTG
1151 CCGTTGCCAT TGCCACCGTA TCGGTACGCC TTTTCTTTAA AAAAACCCAA
1201 TCTTTACAAA GGAACCCGTC ATGA

```

This corresponds to the amino acid sequence [<SEQ ID 802; ORF125-1>] (SEQ ID NO: 802; ORF125-1):

```

1  MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51  AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSVA NMLQLAGWTA
101  VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
151  VSMLLMLLAV LWLSAEVFST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLP
201  LAADYTRHAR RPFATLTAT LAYTLTGCRM YALGLAAALF TGETDVAKIL
251  LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETPVAVGVTL
301  IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFDF
351  AGLVLWLAGF ILYRFLSSG WESSIGLTAP VMSAVAIATV SVRLFPPKKTQ
401  SLQRNPS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 (SEQ ID NO: 800) shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) (SEQ ID NO: 804) from strain A of *N. meningitidis*:

```

5      orf125.pep      10      20      30
                        AGASANNISARFAETPVAVSVTLIGTVLAV
                        ||:||:||:||:||:||:||:||:||:||:||
orf125a      KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAVVGTLAV
                250      260      270      280      290      300

10     orf125.pep      40      50      60
                        MLPVTEYENFLLLLIGSVFAPMGGFDCRLFRLETAX
                        :||:||:||:||:||:||:||:||:||:||:||
orf125a      LLPVTEYENFLLLLIGSVFAPMAAVLIADFFVLKRREEIEG
                310      320      330      340

```

The ORF125a partial nucleotide sequence [<SEQ ID 803>] (SEQ ID NO: 803) is:

```

15      1  ATGTCGGGCA ATGCCTCCTC TCNTTCATCT TCCGCCGCCA TCGGGCTGAT
      51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACACTGC
     101  TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CNGCTCTGCT TTTGGGTCAT
     151  GCCGTCGGCG GCGCGCTGTT TTTTGCGGCG GCGTATATCG GCGCACTGAC
     201  CGGACNCANC TCGATGGAAA GCGTGCGCCT GTCGTTTCGGC AAACGCGGTT
     251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
     301  GTGATGATTT ACGCCGCGC AACGTCAGC TCCGCTTTGG GCAAAGTGTT
     351  GTGGGACGGC GAATCTTTTG TCTGGTGGG ATTGGCAAAC GGCGCGCTGA
     401  TTGTGCTGTG GCTGGTTTTC GCGCACGCA AAACAGGCGG GCTGAAAACC
     451  GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGCTGA GTGCCGAANT
     501  NTTTTCCACG GCAGGCAGCA CCGCCGCANN GGTNNCAGAC GGCATGAGTT
     551  TCGGAACGGC AGTCGAGCTG TCCGCCGTNA TGCCGCTTTC TTGGCTGCCG
     601  CTGGCCGCCG ACTACACGCG CCACGCGCGC CGCCGTTTG CGGCAACCCT
     651  GACGGCAACG CTCGCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG
     701  GTTTGGCAGC GCGTGTGTT ACCGGAGAAA CCGACGTGGC AAAAATCCTG
     751  CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTCG TCCTGTCGAC
     801  CGTTACCACC ACTTTTCTCG ATGCNACTC CGCCGCGGTA AGTGCCAACA
     851  ATATTTCCGC CAAACTTTCG GAAATACCNA TCGCCGTTGC CGTCGCCGTT
     901  GTCGGCACAC TGCTTGCCGT CCTCTGCCC GTTACCGAAT ATGAAAACCT
     951  CCTGCTGCTT ATCGGCTCGG TATTTGCGCC GATGGCGGCG GTTTTGATTG
    1001 CCGACTTTTT CGTCTGAAA CGGCGTGAGG AGATTGAAGG C..

```

This encodes a protein having the partial amino acid sequence [<SEQ ID 804>] (SEQ ID NO: 804):

```

40      1  MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
      51  AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSVA NMLQLAGWTA
     101  VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
     151  VSMLLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP
     201  LAADYTRHAR RPFAATLTAT LAYTLTGCM YALGLAAALF TGETDVAKIL
     251  LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPAVAVAV
     301  VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

```

ORF125a (SEQ ID NO: 804) and ORF125-1 (SEQ ID NO: 802) show 94.5% identity in 347 aa overlap:

-575-

		10	20	30	40	50	60
	orf125a.pep	MSGNASSXSSSA	IGLIWF	GAAV	STAE	ISTG	TLLAP
	orf125-1	MSGNASSPSSSA	IGLIWF	GAAV	STAE	ISTG	TLLAP
5		10	20	30	40	50	60
	orf125a.pep	AYIGALTGX	SMESVR	LSFG	KRGS	VLFS	VANML
	orf125-1	AYIGALTGR	SSMESV	RLSFG	KRGS	VLFS	VANML
10		70	80	90	100	110	120
	orf125a.pep	QAGWTAV	MIYAG	ATVSS	ALGK	VLWD	G
	orf125-1	QAGWTAV	MIYAG	ATVSS	ALGK	VLWD	G
		70	80	90	100	110	120
	orf125a.pep	ESFVW	WALAN	GALIV	LWL	VFG	ARKT
	orf125-1	ESFVW	WALAN	GALIV	LWL	VFG	ARKT
15		130	140	150	160	170	180
	orf125a.pep	GGLKTV	SMLLM	LAVL	WLSA	EXFST	AGSTA
	orf125-1	GGLKTV	SMLLM	LAVL	WLSA	EXFST	AGSTA
		130	140	150	160	170	180
	orf125a.pep	GMSFG	TAVELS	AVMPL	SWLPL	AADY	TRHARR
	orf125-1	GMSFG	TAVELS	AVMPL	SWLPL	AADY	TRHARR
20		190	200	210	220	230	240
	orf125a.pep	PFAATL	TATLAY	TLTG	CWMY	ALGL	AAALF
	orf125-1	PFAATL	TATLAY	TLTG	CWMY	ALGL	AAALF
		190	200	210	220	230	240
	orf125a.pep	TGETD	VAKILL	GAGL	GAAG	ILAV	VLST
	orf125-1	TGETD	VAKILL	GAGL	GAAG	ILAV	VLST
25		250	260	270	280	290	300
	orf125a.pep	VTFTT	FLDAY	SAGV	SANN	ISAK	LSEI
	orf125-1	VTFTT	FLDAY	SAGV	SANN	ISAK	LSEI
		250	260	270	280	290	300
	orf125a.pep	VGTL	LAVLL	PVTE	YENF	LLIG	SVFAP
	orf125-1	IGTV	LAVML	PVTE	YENF	LLIG	SVFAP
30		310	320	330	340	350	360
	orf125a.pep	MAAVLI	ADFF	VLKR	REEIE	G	
	orf125-1	MAAVLI	ADFF	VLKR	REEIE	G	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF125 (SEQ ID NO: 800) shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) (SEQ ID NO: 806) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
35	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
	orf125.pep	MLPVTEYENFLLIGSVFAPM-GGFDCLRFRLTA	64
	orf125ng	MLPVTEYKNFLLIRSVFGPMAGGFDCLRFCLKTA	343
40			

An ORF125ng nucleotide sequence [<SEQ ID 805>] (SEQ ID NO: 805) was predicted to encode a protein having amino acid sequence [<SEQ ID 806>] (SEQ ID NO: 806):

1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA

5
 101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
 151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
 201 PLAADYTRQA RRPFAATLTA TLAYTLTGCW MYALGLAAAL FTGETDVAKI
 251 LLGAGLGITG ILAVVLSTVT TTFLDTYSAG ASANNISARF AEIPVAVGVT
 301 LIRTVLAVML PVTEYKNFLL LIRSVFGPMA GGFDCLFLCL KTA*

Further work revealed the following gonococcal DNA sequence [SEQ ID 807] (SEQ ID NO: 807):

10
 1 ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
 51 TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
 101 TCGCCCCCTT GGGCTGGCAG CGCGGTCTGG CGGCCTGCT TTTGGGTCAT
 151 GCCGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC
 201 CCGACGCAGC TCGATGGAAG GTGTGCGCCT GTCGTTCCGC AAATGCGGTT
 251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
 15
 301 GTGATGATTT ACGTCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
 351 GTGGGACGGC GAATCCTTTG TCTGGTGGG ATTGGCAAAC GGCGCACTGA
 401 TCGTGCTGTG GCTGGTTTTC GGCGCACGCA GAACGGGCGG GCTGAAAACC
 451 GTTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGGTTGA GCGTCGAAGT
 501 GTTCGCTTCG TCCGGCACA ACGCCGCGCC CGCCGTTTCA GACGGCATGA
 20
 551 CCTTCGGAAC GGCAGTCGAA CTGTCCGCCG TCATGCCGCT TTCCTGGCTG
 601 CCGCTGGCCG CCGACTACAC GCGCCAAGCA CGCCGCCCGT TTGCGGCAAC
 651 CCTGACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCCT
 701 TGGGTTTGGC GCGGCTCTG TTTACCGGAG AAACCGACGT GCGGAAAATC
 751 CTGTTGGGCG CGGGCTTGGG CATAACGGGC ATTCTGGCAG TCGTCCTCTC
 25
 801 CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTGCGA
 851 ACAACATTTT CCGCGGTTTT GCGGAAATAC CCGTCGCTGT CCGCGTTACC
 901 CTGATCGGCA CCGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
 951 CTTCCCTGCTG CTTATCGGCT CGGTATTTGC GCCGATGGCG GCGGTTTTGA
 1001 TTGCCGACTT TTTCTGCTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
 30
 1051 TTTGCCGGAC TGGTTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT
 1101 GCTCTCGTCC GGTGGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT
 1151 CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
 1201 CAATCTTTAC AAAGGAACCC GTCATGA

35 This corresponds to the amino acid sequence [SEQ ID 808; ORF125ng-1] (SEQ ID NO: 808; ORF125ng-1):

40
 1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
 101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
 151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
 201 PLAADYTRQA RRPFAATLTA TLAYTLTGCW MYALGLAAAL FTGETDVAKI
 251 LLGAGLGITG ILAVVLSTVT TTFLDTYSAG ASANNISARF AEIPVAVGVT
 301 LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
 351 FAGLVWLWAG FIFYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFfKKT
 45
 401 QSLQRNPS*

ORF125ng-1 (SEQ ID NO: 808) and ORF125-1 (SEQ ID NO: 802) show 95.1% identity in 408 aa overlap:

50
 orf125-1.pep 10 20 30 40 50 60
 MSGNASSPSSSSAIGLIWFGA AVSIAEISTGTLLAPLGWQ RGLAALLLGH AVGGALFFAA
 |||||:||||:|||||

-577-

5	orf125ng-1	MSGNASSPSSSAAIGLVWFGAAVSAEISTGTLLAPLGWQRGLAALLLGHAVGGALFFAA	10	20	30	40	50	60
	orf125-1.pep	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG	70	80	90	100	110	120
10	orf125ng-1	AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG	70	80	90	100	110	120
	orf125-1.pep	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAAQ-VS	130	140	150	160	170	179
15	orf125ng-1	ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSGTNAAPAVS	130	140	150	160	170	180
	orf125-1.pep	DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAAAL	180	190	200	210	220	230
20	orf125ng-1	DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCMYALGLAAAL	180	190	200	210	220	230
	orf125-1.pep	FTGETDVAKILLGAGLGAAGILAVVLSTVTTTTFLDAYSAGASANNISARFAETPVAVGVT	240	250	260	270	280	290
25	orf125ng-1	FTGETDVAKILLGAGLGITGILAVVLSTVTTTTFLDTYSAGASANNISARFAEIPVAVGVT	240	250	260	270	280	290
	orf125-1.pep	LIGTVLAVMLPVTEYENFLLLLIGSVFAPMAAVLIADFFVLKRREEIEGDFAGLVWLWLAG	300	310	320	330	340	350
30	orf125ng-1	LIGTVLAVMLPVTEYKNFLLLLIGSVFAPMAAVLIADFFVLKRREEIEGDFAGLVWLWLAG	300	310	320	330	340	350
	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX	360	370	380	390	400	
35	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX	360	370	380	390	400	

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 809>] (SEQ ID NO: 809):

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCGGGAA GGCTGACCGC
51 GTTGACGCTT GCAGACAAG GTTATCAGAT TGCACTTTC GATAAAGCT
101 GCCGCCGGG CGAACACGCC GCCGCCTATG TAGCCGCCGC CATGCTCGCG

```

5
151 CCTGCAGCGG A.ACGGTCGA AGCCACGCCC GAAGTGGTCA GGCTGGGCAG
201 GCAGAGCATC CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCACA
251 CGATGATGCA GGAAAACGGC AGCCTGATTG TATGGCACGG GCAGGACAAG
301 CCATTATCCA GCGAGTTTCGT CCGCCATCTC AAACGCGGCG GCGT.ACGGA
351 TGACGAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTAA GACGGCATCT ACCTGCCGAC CGAAGC.CAG
451 CTCGACGGGC GGCAATTATA GTCTGCACTT GCCGACGCTT TGGACGAACT
501 GAACGTCCCC TGCCATTGGG AACACGAATG CGTCCCCGAA GCCTGCAAG..

10 This corresponds to the amino acid sequence [<SEQ ID 810; ORF126>] (SEQ ID NO: 810;
ORF126):

15
1 MTRIAILGGG LSGRLTALQL AEQGYQIALF DKSCRRGEHA AAYVAAAML
51 PAAXTVEATP EVVRLGRQSI PLWRGIRCL NTHMMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGXTDDEI VRWRADDIAE REPQLGGRFX DGIYLPTEXQ
151 LDGRQLXSAL ADALDELNVP CHWEHECVPE ACK...

Further work revealed the complete nucleotide sequence [<SEQ ID 811>] (SEQ ID NO: 811):

20
25
30
35
40
1 ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCGGGAA GGCTGACCGC
51 GTTGCACTT GCAGAACAAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101 GCCGCCGGGG CGAACACGCC GCCGCCTATG TTGCCGCCGC CATGCTCGCG
151 CCTGCGGCGG AAGCGGTGCA AGCCACGCCC GAAGTGGTCA GGCTGGGCAG
201 GCAGAGCATC CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCACA
251 CGATGATGCA GGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
301 CCATTATCCA GCGAGTTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACGAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
501 GAACGTCCCC TGCCATTGGG AACACGAATG CGTCCCCGAA GCCTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAAACGCG
601 TGGAACCAAT CCCCCGAGCA CACCAGCACC CTGCGCGGCA TACGCGGCGA
651 AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGTC
701 TGCTCCATCC GCGTTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC
751 TTCGTATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCAGCCAG
801 CGTGCCTTCA GGGTTGGAAC TCTTGTCGCG ACTCTATGCC ATCCACCCCG
851 CTTTCGCGA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
901 CTCAACCACC ACAACCCCGA AATCCGTTAC AACCGCGCCC GACGCCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCCGCCGC CGCCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG
1051 CCCGAACGCG ATAAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA
1101 A

This corresponds to the amino acid sequence [<SEQ ID 812; ORF126-1>] (SEQ ID NO: 812;
ORF126-1):

45
50
1 MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRRGEHA AAYVAAAML
51 PAAEAVEATP EVVRLGRQSI PLWRGIRCL NTHMMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYAKTA
201 WNQSPEHTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHNHPEIRY NRARRLIEIN GLFRHGFMIS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRRQD*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 (SEQ ID NO: 810) shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) (SEQ ID NO: 814) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
5	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP					
	orf126a	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf126.pep	EVVRLGRQSIPLWRGIRCRINHTMMQENGLIVWHGQDKPLSSEFVRHLKRGGXTDDEI					
	orf126a	EVVRLGRQXIPLWRGIRCHLKTTPAMMXENGLIVWHGQDKPLSNEFVRHLKRGGVADDXI					
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE					
	orf126a	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE					
		130	140	150	160	170	180

20 The complete length ORF126a nucleotide sequence [<SEQ ID 813>] (SEQ ID NO: 813) is:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCNGGAA	GGCTGACCGC
	51	ACTGCAGCTT	GCAGAACAAAG	GTTATCAGAT	TGCACTTTTC	GATAAAGGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TTGCCGCCGC	CATGCTCGCG
25	151	CCTGCGGCGG	AAGCGGTCTGA	AGCCACGCCT	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGANCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCATCTG	AAAACGCCTG
	251	CCATGATGCA	NGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAA
	301	CCTTTATCCA	ACGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACNAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
30	451	CTCGACGGGC	GGCAAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	TGCCCCCGAA	GACTTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCGANNA	NACCAGCACC	CTGCGCGGCA	TACGCGGCCA
	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGCC
35	701	TGCTACACCC	GCGCTATCCG	CTNTACATCG	CCCCGAAAGA	AAACCNCGTC
	751	TTCGTCATCG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CACCTGCCAG
	801	CGTGCGTTCC	GGGCTGGAAC	TCTTATCCGC	ACTCTATGCC	GTCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCCACG
	901	CTCAATCACC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCCTGAT
40	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCCGTAA
	1001	CCGCCGCCGC	CGTCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGANGCG
	1051	CCCGAACGCG	ATGAAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

45 This encodes a protein having amino acid sequence [<SEQ ID 814>] (SEQ ID NO: 814):

1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKGCRRGEHA	AAYVAAAMLA
51	PAAEAVEATP	EVVRLGRQXI	PLWRGIRCHL	KTPAMMXENG	SLIVWHGQDK
101	PLSNEFVRHL	KRGGVADDXI	VRWRADDIAE	REPQLGGRFS	DGIYLPTEGQ
151	LDGRQILSAL	ADALDELNVP	CHWEHECAPE	DLQAQYDWLI	DCRGYGAKTA

201 WNQSPXXTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
 301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA
 351 PERDEESGLA YIRRQD*

5 ORF126a (SEQ ID NO: 814) and ORF126-1 (SEQ ID NO: 812) show 95.4% identity in 366 aa overlap:

10	orf126a.pep	10 20 30 40 50 60	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
	orf126-1	10 20 30 40 50 60	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
15	orf126a.pep	70 80 90 100 110 120	EVVRLGRQXIPLWRGIRCHLKTTPAMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDDXI : : :
	orf126-1	70 80 90 100 110 120	EVVRLGRQSIPLWRGIRCRLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI : : :
20	orf126a.pep	130 140 150 160 170 180	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE :
	orf126-1	130 140 150 160 170 180	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE :
25	orf126a.pep	190 200 210 220 230 240	DLQAQYDWLIDCRGYGAKTAWNQSPXXTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP
	orf126-1	190 200 210 220 230 240	GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP
30	orf126a.pep	250 260 270 280 290 300	LYIAPKENXV FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT :
	orf126-1	250 260 270 280 290 300	LYIAPKENHV FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT :
35	orf126a.pep	310 320 330 340 350 360	LNHHNPEIRYNRARRLIEINGLFRHGFMS PAVTAAAVRLAVALFDGKXAPERDEESGLA : : :
	orf126-1	310 320 330 340 350 360	LNHHNPEIRYNRARRLIEINGLFRHGFMS PAVTAAARLAVALFDGKDAPERDKESGLA : : :
40	orf126a.pep		YIRRQDX
	orf126-1		YIRRQDX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 (SEQ ID NO: 810) shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) (SEQ ID NO: 816) from *N.gonorrhoeae*:

orf126.pep MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP 60
 orf126ng MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHAAAYVAAAMLAPAAEAVEATP 60
 5 orf126.pep EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI 120
 orf126ng EVIRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI 120
 orf126.pep VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE 180
 10 orf126ng VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ 180

An ORF126ng nucleotide sequence [[SEQ ID 815](#)] ([SEQ ID NO: 815](#)) was predicted to encode a protein having amino acid sequence [[SEQ ID 816](#)] ([SEQ ID NO: 816](#)):

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA
 51 PAEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
 15 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA
 201 WNQSPEHTST LRGIRGEVRG FTRPKSRSTA PCACCTRAIR STSPRKKTTS
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSM STPPSAKPTS SKWRPGLRPT
 20 301 LNHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

Further work revealed the following gonococcal DNA sequence [[SEQ ID 817](#)] ([SEQ ID NO: 817](#)):

1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC
 25 51 ATTGCAGCTT GCAGAACAAG GTTATCAGAT TGAACTTTTC GACAAGGGCA
 101 CCGGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG
 151 CCTGCGGCGG AAGCGGTCGA GGCAACGCCC GAAGTCATCA GGCTGGGCAG
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA
 251 CGATGATGCA GGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
 30 301 CCATTATCCA GCGAGTTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
 351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCCG
 401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
 451 CTCGACGGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
 501 GAACGTCCCT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG
 35 551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG
 601 TGGAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA
 651 AGTGGCGCGG GTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC
 751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCAGCCAG
 40 801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG
 851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG
 901 CTCAACCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT
 951 CGAAATCAAC GGCTTTTTC GGCACGGCTT TATGATTTCC CCCGCCGTAA
 1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACCGG
 45 1051 CCCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA
 1101 A

This corresponds to the amino acid sequence [[SEQ ID 818; ORF126ng-1](#)] ([SEQ ID NO: 818; ORF126ng-1](#)):

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA
 51 PAEEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA
 201 WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

10 ORF126ng-1 (SEQ ID NO: 818) and ORF126-1 (SEQ ID NO: 812) show 95.1% identity in 366 aa overlap:

		10	20	30	40	50	60
	orf126-1.pep	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKGCRGEHA	AAYVAAAMLA	PAEEAVEATP
		:					
15	orf126ng-1	MTRIAVLGGG	LSGRLTALQL	AEQGYQIELF	DKGTRQGEHA	AAYVAAAMLA	PAEEAVEATP
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf126-1.pep	EVVRLGRQSI	PLWRGIRCRL	NLTHTMMQENG	SLIVWHGQDK	PLSSEFVRHL	KRGGVADDEI
		:					
20	orf126ng-1	EVIRLGRQSI	PLWRGIRCRL	NTLTMMQENG	SLIVWHGQDK	PLSSEFVRHL	KRGGVADDEI
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf126-1.pep	VRWRADDIAE	REPQLGGRFS	DGIYLPTEGQ	LDGRQILSAL	ADALDELNVP	CHWEHECVPE
		:					
25	orf126ng-1	VRWRADEIAE	REPQLGGRFS	DGIYLPTEGQ	LDGRQILSAL	ADALDELNVP	CHWEHECAPQ
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf126-1.pep	GLQAQYDWL	IDCRGYGAKT	AWNQSPEHTS	TLRGIRGEVAR	VYTPEITLNR	PVRLHPRYP
		:					
30	orf126ng-1	DLQAQYDWV	IDCRGYGAKT	AWNQSPEHTS	TLRGIRGEVAR	VYTPEITLNR	PVRLHPRYP
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHV	FVIGATQIES	ESQAPASVRS	GLELLSALYA	IHPAFGEADI	LEIATGLRPT
		:					
35	orf126ng-1	LYIAPKENHV	FVIGATQIES	ESQAPASVRS	GLELLSALYA	VHPAFGEADI	LEIAAGLRPT
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf126-1.pep	LNHHNPEIRY	NRARRLIEIN	GLFRHGFMS	PAVTAAARLA	VALFDGKDAP	ERDKESGLA
		:					
40	orf126ng-1	LNHHNPEIRY	SRERRLIEIN	GLFRHGFMS	PAVTAAAVRL	AVALFDGKDAP	ERDEESGLA
		310	320	330	340	350	360
		YIRRQDX					
	orf126-1.pep	YIRRQDX					
45	orf126ng-1	YIGRQDX					

Furthermore, ORF126ng-1 (SEQ ID NO: 818) shows homology to a putative *Rhizobium* oxidase flavoprotein (SEQ ID NO: 1163):

gi|2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
Length = 327

Score = 169 bits (423), Expect = 3e-41

Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)

5 Query: 3 RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXXXXX 62
RI V G G++G A QL G+++ L ++ G
Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEPEV 60

10 Query: 63 IRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
+ LGR + W + G+L+V G+D F R G DE+
Sbjct: 61 LTLGRLAADWEEA-----LPGHVHRRGTLLVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113

Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
IA EP L GRF ++ E LD RQ L+ALA L++ + +
Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165

15 Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYPLY 242
+D V+DC G LRG+RGE+ V T E++L+RPVRLHPR+P+Y
Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVGEMLCVETTEVSLSRPVRLHPRHPIY 218

Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
I P++ + F++GAT IES+ P + RS +ELL+A YA+HPAFGEA + E AG+RP
Sbjct: 219 IVPRDKNRFMVGATMIESDDGGPITARSLMELLNAAAYAMHPAFGEARVTETGAGVRPAYP 278

20 Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMISP 331
+ P R ++E R + +NGL+RHGF+++P
Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305

25 This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* [<SEQ ID 819>] (SEQ ID NO: 819):

30 1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTAA TCTGCAGAAT GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGTCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
35 301 AAGGCGGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAGATGAA
351 TGAAAATCTA GTAACCTTTA aTTTGCAAGA AGTCCGCCAG TTCGTGTAGT
401 GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
451 GTAG

40 This corresponds to the amino acid sequence [<SEQ ID 820; ORF127>] (SEQ ID NO: 820; ORF127):

1 MTDNRGFTLV ELISVVLILS VLALIVPSY RNYVEKAKIN AVRAALLENA

51 HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
 101 KAVAIDKDKN PFIKMNENL VTFICKKSAS SCSDGLDYFK GNDKCKLLK
 151 *

5 Further work revealed the following DNA sequence [<SEQ ID 821>] (SEQ ID NO: 821):

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
 51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
 101 TTGAGAAAGC AAAGATAAAT GCAGTGC GGG CAGCCTTGTT AGAAAATGCA
 151 CATT TTTATGG AAAAGTTT TCTGCAGAAT GGGAGGTTTA AACAAACATC
 10 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
 201 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
 251 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
 301 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTTCG TGTAGTGACG
 351 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG
 15

This corresponds to the amino acid sequence [<SEQ ID 822; ORF127-1>] (SEQ ID NO: 822; ORF127-1):

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA
 51 HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
 20 101 AVAIDKDKNP FIIKMNENLV TFICKKSASS CSDGLDYFKG NDKCKLLK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 (SEQ ID NO: 820) shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) (SEQ ID NO: 824) from strain A of *N. meningitidis*:

30 orf127.pep 10 20 30 40 50 60
 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLENA HFMEKFYLQN
 orf127a MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLENA HFMEKFYLQN
 35 orf127.pep 70 80 90 100 110 120
 GRFKQTSTKW PSLPIKEAEG FCIRLNGI VARXALDSKFML KAVAIDKDKNP FIIKMNENL
 orf127a GRFKQTSTKW PSLPIKEAEG FCIRLNGI -ARGALDSKFML KAVAIDKDKNP FIIKMNENL
 40 orf127.pep 130 140 150
 VTFICKKSASS CSDGLDYFKG NDKCKLLKX
 orf127a VTFICKKSASS CSDGLDYFKG NDKCKLLKX
 120 130 140 150

The complete length ORF127a nucleotide sequence [<SEQ ID 823>] (SEQ ID NO: 823) is:

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT

-585-

51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
 101 TTGAGAAAGC AAAGATAAAT ACAGTGCGGG CAGCCTTGTT AGAAAATGCA
 151 CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GGGAGATTTA AACAAACATC
 201 TACCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
 251 GTTTGAATGG AATCGCGCGC GGGGCCTTAG ACAGTAAATT CATGTTGAAG
 301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
 351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
 401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

10 This encodes a protein having amino acid sequence [<SEQ ID 824>] (SEQ ID NO: 824):

1 MTDNRGFTLV ELISVVLILS VLALIVPSY RNYVEKAKIN TVRAALLEN
 51 HFMEKFYLQN GRFKQSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
 101 AVAIDKDKNP FIIKMENLV TFICKKSASS CSDGLDYFKG NDKDCKLLK*

15 ORF127a (SEQ ID NO: 824) and ORF127-1 (SEQ ID NO: 822) show 99.3% identity in 149 aa overlap:

		10	20	30	40	50	60
orf127a.pep	MTDNRGFTLV	ELISVVLILS	VLALIVPSY	RNYVEKAKIN	TVRAALLEN	HFMEKFYLQN	
20 orf127-1	MTDNRGFTLV	ELISVVLILS	VLALIVPSY	RNYVEKAKIN	AVRAALLEN	HFMEKFYLQN	
		10	20	30	40	50	60
		70	80	90	100	110	120
orf127a.pep	GRFKQSTKW	PSLPIKEAEG	FCIRLNGIAR	GALDSKFMLK	KAVAIDKDKNP	FIIKMENLV	
25 orf127-1	GRFKQSTKW	PSLPIKEAEG	FCIRLNGIAR	GALDSKFMLK	KAVAIDKDKNP	FIIKMENLV	
		70	80	90	100	110	120
		130	140	150			
orf127a.pep	TFICKKSASS	CSDGLDYFKG	NDKDCKLLKX				
30 orf127-1	TFICKKSASS	CSDGLDYFKG	NDKDCKLLKX				
		130	140	150			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 (SEQ ID NO: 820) shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) (SEQ ID NO: 826) from *N.gonorrhoeae*:

35	orf127.pep	MTDNRGFTLV	ELISVVLILS	VLALIVPSY	RNYVEKAKIN	AVRAALLEN	HFMEKFYLQN	60
	orf127ng	MTDNRGFTLV	ELISVVLILS	VLALIVPSY	RNYVEKAKIN	AVRAAFLEN	HFMEKFYLQN	60
	orf127.pep	GRFKQSTKW	PSLPIKEAEG	FCIRLNGIAR	XALDSKFMLK	KAVAIDKDKNP	FIIKMENL	120
40	orf127ng	GRFKQSTKW	PSLPIKEAEG	FCIRLNGI -	ARGALDSKFMLK	KAVAIDKDKNP	FIIKMENL	119
	orf127.pep	VTFICKKSASS	CSDGLDYFKG	NDKDCKLLK				150
	orf127ng	VTFICKKSASS	CDRLDYFKG	NDKDCKLLK				149

The complete length ORF127ng nucleotide sequence [<SEQ ID 825>] (SEQ ID NO: 825) is:

```

1  ATGACTGATA ATCGGGGGTT TACACTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
5  151 CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGATTTA AACAAACATC
201 TACCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
10 401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence [<SEQ ID 826>] (SEQ ID NO: 826):

```

1  MTDNRGFTLV ELISVVLILS VLALIVPSY RNYVEKAKIN AVRAAFLENA
15 51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIIKMENLV TFICKKSASS CSDRLDYFKG NDKDCKLLK*

```

ORF127ng (SEQ ID NO: 826) and ORF127-1 (SEQ ID NO: 822) show 100.0% identity in 149 aa overlap:

```

20      10      20      30      40      50      60
orf127-1.pep MTDNRGFTLV ELISVVLILS VLALIVPSY RNYVEKAKIN AVRAALLEN AHFMEKFYLN
|
orf127ng-1   MTDNRGFTLV ELISVVLILS VLALIVPSY RNYVEKAKIN AVRAALLEN AHFMEKFYLN
|
      10      20      30      40      50      60

25      70      80      90     100     110     120
orf127-1.pep GRFKQTSTKW PSLPIKEAEG FCIRLNGI ARGALDSKFMLK AVAIDKDKNP FIIKMENLV
|
orf127ng-1   GRFKQTSTKW PSLPIKEAEG FCIRLNGI ARGALDSKFMLK AVAIDKDKNP FIIKMENLV
|
      70      80      90     100     110     120

30      130     140     150
orf127-1.pep TFICKKSASS CSDGLDYFKG NDKDCKLLKX
|
orf127ng-1   TFICKKSASS CSDGLDYFKG NDKDCKLLKX
|
      130     140     150

```

35 This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 98

40 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 827>] (SEQ ID NO: 827)

-587-

1 ..GTGTCGCTGG CTTCGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
 51 CAACCAAATG CGGAAAACCG GTGGAGCTAT CTGCGGTTTT CTGTGTCCAAT
 101 ATTTATCTGG GGTTCAGCA GGGGTATTTT GATTTGAGTG CCGACGAGAA
 151 CCCCCTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
 5 201 TGTATCCCTT TTTGCTGATA TTTTGTGCA AAAAAACCAA ATCGCTACGG
 251 GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
 301 GTTTTTGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCCAATACTT
 351 ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTTGGCAGG TTCGCTGCTG
 401 GCGGTTTACG GGCAAACGCA AAACGGCAGA CGGCAAACAG CAAATGGAAA
 10 451 ACGCAGTTG CTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
 501 TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
 551 CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
 601 TCCGACCCGC ATCCTGTCGG CAAGCCCAT CGTATTTGTC GGCAAAATCT
 651 CTTATTCCCT ATACCTGTAC CATTGGATT TTATTGCTTT CGCTCCGCTC
 15 701 ATTAGAGGCG GGAAACAGCT CGGACTGCCT GCCG...

This corresponds to the amino acid sequence [<SEQ ID 828; ORF128>] (SEQ ID NO: 828; ORF128):

1 ..VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGFAQ GYFDLSADEN
 20 51 PVLHIWSLAV EEQYLLYPL LLIFCCKKTK SLRVLRNISI ILFLILTASS
 101 FLPSGFYTDI LNQPNTYYLS TLRFPPELLAG SLLAVYGQTQ NGRRTANGK
 151 RQLLSSLCFG ALLACLFVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
 201 PTRILSASPI VFVGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..

25 Further work revealed the complete nucleotide sequence [<SEQ ID 829>] (SEQ ID NO: 829):

1 ATGCAAGCTG TCCGATACAG ACCGGAAATT GACGGATTGC GGGCCGTCGC
 51 CGTGCTATCC GTCATGATT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
 101 GATTCCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
 151 GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
 30 201 TTATACCCGC AGGATTAAGC GGATTATACC TGCCTTTATT GCGGCCGTGT
 251 CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTCAAC
 301 CAAATGCGGA AAACCGTGGA GCTTCTGCG GTTTTCTTGT CCAATATTTA
 351 TCTGGGGTTT CAGCAGGGGT ATTTGATTT GAGTGCCGAC GAGAACCCCG
 401 TACTGCATAT CTGCTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
 35 451 CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT
 501 GCGTAACATC AGCATCATCC TGTTTTGTAT TTTGACTGCC TCATCGTTTT
 551 TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCTA TACTTATTAC
 601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTTCG TGCTGGCGGT
 651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC
 40 701 AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTGCCTG CCTGTTCTGT
 751 ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG
 801 CCTGCTGACG GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA
 851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
 901 TCCCTATACC TGTACCATTT GATTTTATT GCTTTCGCCC ATTACATTAC
 45 951 AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCCGGTTGA
 1001 CGGCCGGAAT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA
 1051 AAACGGAAGA TGACCTCAA AAAGGCATT TTCTGCCTCT ATCTCGCCCC
 1101 GTCCCTGATA CTGTGCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC
 1151 AGGAACACCT CCGCCCGTTG CCCGGCGCGC CCCTTGCTGC GGAAAATCAT
 50 1201 TTTCCGAAA CCGTCTGAC CCTCGCGGAC TCGCACGCCG GACACCTGAG
 1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGTTGGAAA GCCAAAATCC
 1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
 1351 AATCCGTTAT GTCGAAAATA CCGGATGAA GTTGAAAAAG CCGAAGCCGT
 1401 TTTCTTGCC CAATCTATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCGA
 55 1451 GATTTGAAGC GCAATCCTTC CTAATACCCG GGTTCACGAC CCGATTACGG
 1501 GAAACCGTCA AAAGGATAGC CGCCGTCAA CCCGTCTATG TTTTGTGAAA

1551 CAACACATCA ATCAGCCGTT CGCCCCTGAG GGAGGAAAAA TTGAAAAGAT
 1601 TTGCCGCAAA CCAATATCTC CGCCCCATTG AGGCTATGGG CGACATCGGC
 1651 AAGAGCAATC AGGCGGTCTT TGATTTGATT AAAGATATTC CCAATGTGCA
 1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
 1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
 1801 TATATGGGGC GGAATTCCA CAAACACGAA CGCCTGCTTA AATCTTCCCA
 1851 CGGCGGCGCA TTGCAGTAG

This corresponds to the amino acid sequence [<SEQ ID 830; ORF128-1>] (SEQ ID NO: 830;

ORF128-1):

1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT
 51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
 101 QMRKTVELSA VFLSNIYLG F QQGYFDLSAD ENPVLHIWSL AVEEQYYLLY
 151 PLLLIFFCCKK TKSLRVLRNI SIILFLILTA SSFLPSGFYT DILNQPNNTYY
 201 LSTLRFPELL AGSLLAVYGQ TQNGRRQTAN GKRQLLSSLC FGALLACLFV
 251 IDKHNPFIPIG MTL LLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
 301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSL SYYLIEQPLR
 351 KRKMTFFKAF FCLYLAPSLI LVGYNLYARG ILKQEHLRPL PGAPLAAENH
 401 FPETVLTGLD SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
 451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFPARFR
 501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPNVHWVD AQKYL PKNTV EIYGRYLYGD QDHLTYFGSY
 601 YMGREFHKHE RLLKSSHGGA LQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical integral membrane protein HI0392 of *H. influenzae* (accession number U32723) (SEQ ID NO: 1164)

ORF128 (SEQ ID NO: 828) and HI0392 (SEQ ID NO: 1164) show 52% aa identity in 180aa overlap:

Orf128: 1 VSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG FQQGYFDLSADENPVLHIWSLAV 60
 ++L S IAS IF+Y DFN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWSLAV
 HI0392: 46 MALVSFIASAI F IYND FNKL RKTIELAIAFLSNFYLG L TQGYFDLSANENPVLHIWSLAV 105
 Orf128: 61 EEQXXXXXXXXXIFCCKKTKSLRVLRNISIILFLILTASSFLPSGFYTDILNQPNNTYYLS 120
 E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS
 HI0392: 106 EGQYYLIFPLILILAYKKFREV KVLFIITLILFFILLATSFVSANFYKEVLHQPNIIYYLS 165
 Orf128: 121 TLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLCFGALLACLFVIDKHNPFIPIGMT 180
 LRFPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T
 HI0392: 166 NLRFPPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF128 (SEQ ID NO: 828) shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) (SEQ ID NO: 832) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF128a nucleotide sequence [<SEQ ID 831>] (SEQ ID NO: 831) is:

30	1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCGTTCGC
	51	CGTGCTATCC	GTCATGATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
	151	GGCATCATTC	TTTCTGAAAT	ACAGAACGGT	TCCTTTTCTT	TCCGGGATTT
35	201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	CGCGCCGTGT
	251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
	301	CAAATGCGGA	AAACCGTGGG	GCTTTCTGCG	GTTTTCTTGT	CCAATATTTA
	351	TCTGGGGTTT	CAGCAGGGGT	ATTTTCGATT	GAGTGCCGAC	GAGAACCCCG
40	401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
	451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAAATCG	TACGGGTGCT
	501	GGCTAACATC	AGCATCATCC	TATTTCTGAT	TTTGACTGCC	ACATCGTTTT
	551	TGCCAAGCGG	GTTTTATACC	GATATTTCTA	ACCAACCCAA	TACTATTATC
45	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTTCG	TGCTGGCGGT
	651	TTACGGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAAACGGC
	701	AGTTGCTTTT	ATCACTCTGC	TTCGGCGCAT	TGCTTGCCGT	CCTGTTCGTG
	751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCCTGG	TCCTTCCCTG
50	801	CCTGCTGACG	GCATCGCTTA	TCCGGAGTAT	GCAATCCGGG	AACTTCCGA
	851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTTCGGCA	AATCTCTTAT
	901	TCCCTATACC	TGTACCATTG	GATTTTTTATT	GCTTTCGCCC	ATTACATTAC
	951	AGGCGACAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
50	1001	CGGCCCGATT	TTCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
	1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
	1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAAC
	1151	AGGAACACCT	CCGCCCCGTT	CCCGGCGCGC	CCCTTGCTGC	GGAAAAATCAT

-590-

5
10
15

```

1201 TTTCCGGA   CCGTCCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCC
1251 GGGGTTTCTG GATTATGTCG GCAGCCGGA  AGGGTGGAAA GCCAAAATCC
1301 TGTCCTCGA  TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
1351 AACCCGTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT
1401 TTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
1451 GATTGAAGC  GCAATCCTTC CTAATACCCG GGTCCCAGC  CCGATTGAGG
1501 GAAACCGTCA AAAGGATAGC CGCCGTCAA  CCCGTCTATG TTTTGTCAA
1551 CAACACATCA ATCAGCCGTT CGCCCCTGAG GGAGGAAAA  TTGAAAAGAT
1601 TTGCCGCAA  CCAATATCTC CGCCCCATTC AGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGGCGGTCTT TGATTTGATT AAAGATATTC CCAATGTGCA
1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGGC GGAATTTCA  CAAACACGAA CGCCTGCTTA AATCTTCTCG
1851 CGACGGCGCA TTGCAGTAG

```

This encodes a protein having amino acid sequence [<SEQ ID 832>] (SEQ ID NO: 832):

20
25
30

```

1  MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT
51  GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTVELSA VFLSNIYLG F QQGYFDLSAD ENPVLHIWSL AVEEQYLLY
151 PLLLIFCCKK TKSLRVLRNI SIILFLILTA TSFLPSGFYT DILNQPNYY
201 LSTLRFPELL AGSLLAVYGQ TQNGRRQTAN GKRQLSSLC FGALLACLFV
251 IDKHNPFIPG MTL LLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR
351 KRKMTFKKAF FCLYLAPSLI LVGYNLYARG ILKQEHLRPL PGAPLAAENH
401 FPETVLTLDG SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCKRYRDE VEKAEAVFIA QFYDLRMGGQ PVPFQAQSF LIPGFPARFR
501 ETVKRIAARK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
551 KSNQAVFDLI KDIPNVHWVD AQKYLPKNTV EIYGRYLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKSSRDGA LQ*

```

ORF128a (SEQ ID NO: 832) and ORF128-1 (SEQ ID NO: 830) show 99.5% identity in 622 aa overlap:

35
40
45
50

```

orf128a.pep  MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGV DIFVISGFLITGIILSEIQNG
orf128-1     MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGV DIFVISGFLITGIILSEIQNG

orf128a.pep  SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG F
orf128-1     SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG F

orf128a.pep  QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA
orf128-1     QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA

orf128a.pep  TSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC
orf128-1     SSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC

orf128a.pep  FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY
orf128-1     FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFVGKISY

orf128a.pep  SLYLYHWIFI AFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
orf128-1     SLYLYHWIFI AFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF

```

-591-

	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
5	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFILPGFPARFRETVKRIA AVKPVYVFANNTSISRSPREEKLKRFAANQYL
	orf128-1	PVPRFEAQSFILPGFPARFRETVKRIA AVKPVYVFANNTSISRSPREEKLKRFAANQYL
10	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
15	orf128-1	YMGREFHKHERLLKSSSHGGALQX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF128 (SEQ ID NO: 828) shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) (SEQ ID NO: 834) from *N. gonorrhoeae*:

20	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
	orf128.pep	LSNIYLGFGQGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCCKKTKSLRVLRNISI	90
	orf128ng	LSNIYLGFR LGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCYKTKTKSLRVLRNISI	172
25	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSL LAVYGQTQNGRRQTENGK	232
	orf128.pep	RQLLSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
30	orf128ng	RQLLSSLCFGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292
	orf128.pep	VFVGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA	244
	orf128ng	VFVGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR	352

35 The complete length ORF128ng nucleotide sequence [<SEQ ID 833>] (SEQ ID NO: 833) is:

	1	ATGCAAGCTG	TCCGATACAG	GCCTGAAATT	GACGGATTGC	GGGCCGTCGC
	51	CGTGCTATCC	GTCATTATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCGGGATT	CCTCATTACC
	151	AACATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTT	TCCGGGATTT
40	201	TTATACCGCG	AGGATTAAGC	GGATTATCC	TGCTTTTATT	GCGGCCGTGT
	251	CCCTGGCTTC	GGTGATTGCT	TCTCAAATCT	TCCTTTACGA	AGATTTC AAC
	301	CAAATGAGGA	AAACCATAGA	GCTTTCTACG	GTTTTTTTGT	CCAATATTTA
	351	TTTGGGGTTC	CGATTGGGGT	ATTTTCGATT	GAGTGCCGAC	GAGAACCCCG

5
10
15
20
25
30

```

401 TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT
451 CCTCTTTTGC TGATATTCTG TTACAAAAAA ACCAAATCAC TACGGGTGCT
501 GCGTAATATC AGCATCATCC TGTTCCTGAT TTGACCGCA TCATCGTTTT
551 TGCCGCGCCG GTTTTATACC GACATCCTCA ACCAACCcaa TACTTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTCGC TGTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAAT GGAAAACGGC
701 AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTGTCTG CCTGTTTCGTG
751 ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCCTGC TCCTTCCCTG
801 CCTGCTGACG GCGCTGCTTA TCCGGAGTAT GCAATACGGG ACACCTCCGA
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTG GATTTTATT GCCTTCGCCC ATTACATTAC
951 AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA
1001 CGGCCGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTAGA
1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTTT ATCTCGCCCC
1101 GTCCCTGATG CTTGTCGGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCGCTG CCCGGCACGC CCGTGTCTGC GGAAAATAAT
1201 TTTCCGAAA CCGTCTTGAC CCTCGCGAC TCGCACGCCG GACACCTCGC
1251 GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGGTGGA GCTAAAATCC
1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TGGATGAGAA GCTGGCAGAC
1351 AACCCGTTGT GCCGAAAATA CCGGATGAA GTTGAAAAAG CCGAAGCTGT
1401 TTTCAATGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
1451 GATTTGAAGC GCAATCCTTC CTGATACCCG GGTCAAAGC CCGATTCAAG
1501 GAAACCGTCA AGAGGATAGC CGCCGTCAA CCTGTATATG TTTTGTCAA
1551 CAATACATCA ATCAGCCGTT CTCCCTTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCTATAAA CCAATACCTC CGGCCTATTC GGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGGCGGTCTT TGATTGGTT AAAGATATTC CCAATGTGCA
1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG
1751 GACGTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGGC GGGAAATTCA CAAACACGAA CGCCTGCTCA AGCATTCCCG
1851 AGGCGGCGCA TTGCAGTAG

```

This encodes a protein having amino acid sequence [SEQ ID 834] (SEQ ID NO: 834):

35
40
45

```

1 MQAVRYRPEI DGLRAVAVLS VIIFHLNNRW LPGGFLGVDI FFVISGFLIT
51 NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTIELST VFLSNIYLG FRLGYFDLSAD ENPVLHIWSL AVEEQYYLLY
151 PLLLIFYCYK TKSLRVL RNI SIILFLILTA SSFLPAGFYT DILNQPNTRY
201 LSTLRFPELL VGSLLAVYGQ TQNGRRQTEN GKRQLSLLC FGALLVCLFV
251 IDKHPFPIG ITLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLI SYYLIEQPLR
351 KRKMTFKKAF FCLYLAPSLM LVGYNLYSRG ILKQEHLRPL PGTPVAAENN
401 FPETVLTLD SHAGHLRGFL DYVGREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPFEAQSF LIPGFKARFR
501 ETVKRIA AVK PVYVFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDIG
551 KSNQAVFDLV KDIPNVHVD AQKYLKNTV EIHGRYLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKHSRGG A LQ*

```

ORF128ng (SEQ ID NO: 834) and ORF128-1 (SEQ ID NO: 830) show 95.7% identity in 622 aa overlap:

50

```

orf128-1.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
| | | | | : | | | | | : | | | | |
orf128ng MQAVRYRPEIDGLRAVAVLSVIIFHLNNRWLPGGFLGVDIFFVISGFLITNIILSEIQNG

orf128-1.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
| | | | | : | | | | | : | | | | |
orf128ng SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG

```


	orf128-1.pep	orf128ng
5	orf128-1.pep	SSFLPSPGFYTDILNQPNITYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
	orf128ng	SSFLPAGFYTDILNQPNITYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGKRQLLSSLC
	orf128-1.pep	FGALLACL FVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVGKISY
	orf128ng	FGALLVCL FVIDKHDPFI PGITLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVGKISY
10	orf128-1.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFFKAF
	orf128ng	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFFKAF
	orf128-1.pep	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128ng	FCLYLAPSLMLVGYNLYSRGILKQEHLRPLPGTPVAAENHFPETVLTGLGDSHAGHLRGFL
15	orf128-1.pep	DYVGSREGWKAKIILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128ng	DYVGGREGWKAKIILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128-1.pep	PVPRFEAQSF LIPGFPARFRET VKRIA AVKPVYVFANNTSISRSP LREEKLKRFAANQYL
	orf128ng	PVPRFEAQSF LIPGFKARFRET VKRIA AVKPVYVFANNTSISRSP LREEKLKRFAINQYL
20	orf128-1.pep	RPIQAMGDIGKSNQAVFDLIK DIPNVHWVDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
	orf128ng	RPIRAMGDIGKSNQAVFDLVKDIPNVHWVDAQYLPKNTVEIHGRYLYGDQDHLTYFGSY
	orf128-1.pep	YMGREFHKHERLLKSSHGGALQX
	orf128ng	YMGREFHKHERLLKHSRGGALQX

30 In addition, [ORF218ng] ORF128ng (SEQ ID NO: 834) shows homology to a hypothetical
H.influenzae protein (SEQ ID NO: 1164):

```

35  sp|P43993|Y392_HAEIN HYPOTHETICAL PROTEIN HI0392 |gi|1074385|pir|B64007
    hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
    |gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
    influenzae] Length = 245
    Score = 239 bits (604), Expect = 3e-62
    Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

    Query: 38  VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
               +DIFFVISGFLIT II++EIQ  SFS + FYTRRIKRIYP                      F+Y
40  Sbjct: 1    MDIFFVISGFLITGIIITEIQNSFSLKQFYTRRIKRIYPAFITVMALVSFIASAIIFIYN 60

    Query: 98  DFNQMRKTIELSTVFLSNIIYLGFRLLGYFDLSADENPVLHIWLSLAVEEQXXXXXXXXXXIFC 157
               DFN++RKTIEL+  FLSN YLG  GYFDLSA+ENPVLHIWLSLAVE Q              I
    Sbjct: 61  DFNKLARKTIELAIAFLSNFYLGLTQGYFDLSANENPVLHIWLSLAVEGQYYLIFPLILILA 120

    Query: 158 YKKTKSLRVLNRNISIIILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSL LAV 217
               YKK + ++VL  I++ILF IL A+SF+ A FY ++L+OPN YYLS LRFPELLVGSL LA+

```

Sbjct: 121 YKKFREVKVLFIITLILFFILLATSFVVSANFYKEVLHQPNIIYLSNLRFPPELLVGSLLAI 180

Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDPFIPGIT 262

Y N + Q +L++L L CLF+++ + FIPGIT

Sbjct: 181 YHNSLN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

5 This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

10 The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 835>] (SEQ ID NO: 835):

15
 1 ..ATTATTTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
 51 GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT
 101 TGGCGTTGGC GCGCCTGATT CACTTGAAA AAGCCGGTGC GCCGATGCGC
 151 GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
 201 CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTGCGGCG TATGTGTGGT
 251 TTCCGTTTTT CGTC..

20 This corresponds to the amino acid sequence [<SEQ ID 836; ORF129>] (SEQ ID NO: 836; ORF129):

1 ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
 51 VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFV..

Further work revealed the complete nucleotide sequence [<SEQ ID 837>] (SEQ ID NO: 837):

25
 1 ATGGATTTC GTTTGACAT TATTACGAA TACCGCTGGA TGTTCCTTTA
 51 CCGCGCACTG ACGACCTTGG GGCTGACGGT CGTGCAACG GCGGGCGGTT
 101 CCGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTGGAAAAA
 151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
 201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTGTG CAGATTGTGA
 30 251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
 301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
 351 GATTGCCGGT TCTTTGGCAC TGATCGCAA CTCGGGGGCG TATATCTGTG
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
 451 GCGCGTTCTT TGGGGCTGAC CTATCCGCG GCGATGCGCT ATGTGATTCT
 35 501 GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG
 601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTAGGCT
 40 701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

This corresponds to the amino acid sequence [<SEQ ID 838; ORF129-1>] (SEQ ID NO: 838; ORF129-1):

1 MDRFDIIE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK

-595-

51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

5

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 (SEQ ID NO: 836) shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) (SEQ ID NO: 840) from strain A of *N. meningitidis*:

10 orf129.pep I IYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 orf129a MDRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 15 orf129.pep ALRKVSLLYVTTLFRGTPLFVQIVIWAYVWFPPFFV
 orf129a ALRKVSLLYVTTLFRGTPLFVQIVIWAYVWFPPFFVHPSDGILVSGEAAIALRRYGPLIAG
 20 orf129a SLALIANSQA YICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS

The complete length ORF129a nucleotide sequence [<SEQ ID 839>] (SEQ ID NO: 839) is:

25 1 ATGGATTTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTTCTTTA
 51 CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT
 101 CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAGAAA
 151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
 201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTGTG CAGATTGTGA
 251 TTTGGGCGTA TGTGTGTTT CCGTTTTCG TCCATCCTTC AGACGGCATT
 30 301 TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
 451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
 501 GCCGAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
 35 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGAGATTG
 601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTAGGCT
 701 GGATATTCCT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

40 This encodes a protein having amino acid sequence [<SEQ ID 840>] (SEQ ID NO: 840):

45 1 MDRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
 51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

ORF129a (SEQ ID NO: 840) and ORF129-1 (SEQ ID NO: 838) show 100.0% identity in 248 aa overlap:

```

5  orf129a.pep  MDFRFDIIYEYRWFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
   orf129-1    MDFRFDIIYEYRWFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW

   orf129a.pep  ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRGYGPLIAG
   orf129-1    ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRGYGPLIAG

10  orf129a.pep  SLALIANS GAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS
   orf129-1    SLALIANS GAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS

   orf129a.pep  EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
   orf129-1    EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE

15  orf129a.pep  KRYNPQHRX
   orf129-1    KRYNPQHRX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 (SEQ ID NO: 836) shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) (SEQ ID NO: 842) from *N.gonorrhoeae*:

```

   orf129.pep      I IYEYRWFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW      54
   orf129ng        MDFRFDIIYEYRWFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW      60

   orf129.pep      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFV                          88
   orf129ng        ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVILHTAFLGNAMRQSRRVPDKGRWIAG      120

```

An ORF129ng nucleotide sequence [<SEQ ID 841>] (SEQ ID NO: 841) was predicted to encode a protein having amino acid sequence [<SEQ ID 842>] (SEQ ID NO: 842):

```

30      1  MDFRFDIIYE YRWFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
      51  AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PPFVILHTAF
      101  LGNAMRQSRR VPDKGRWIAG SLELNCQPRG RKTRGEFPPG ESNLGTGEPN
      151  PLSMGQRRFP GCENWYPPQN FIKK*

```

Further work revealed the following gonococcal sequence [<SEQ ID 843>] (SEQ ID NO: 843):

```

40      1  ATGGATTTTc gtTTTGACAT TATTAcgaA TACCGCTGGA TGTTTCTTTA
      51  CGGCGCACTG Acgaccttgg ggctgacggt cgtggcgacg gCGGCGGTT
      101  CGGtattggG TCTGTTGTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
      151  GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
      201  GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCTGTTGTG CAGATTGTGA

```

5
10

```

251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451 GCGTGTCTTT TGGGACTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG
601 GCGTATGTTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651 GCTTTACACC GCCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701 GGATATTCTT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence [SEQ ID 844; ORF129ng-1] (SEQ ID NO: 844; ORF129ng-1):

15

```

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGPIIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
201 AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

20 ORF129ng-1 (SEQ ID NO: 844) and ORF129-1 (SEQ ID NO: 838) show 99.2% identity in 248 aa overlap:

25

```

orf129-1.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
|
orf129ng-1 MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
|
orf129-1.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRYGPIIAG
|
orf129ng-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRYGPIIAG
|
orf129-1.pep SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS
|
orf129ng-1 SLALIANSQA YICEIFRAGI QSIDKGQMEA ACSLGLTYPQ AMRYVILPQA LRRMLPPLAS
|
orf129-1.pep EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE
|
orf129ng-1 EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT AALIYLLMTT FLGWIFLRLE
|
orf129-1.pep KRYNPQHRX
|
orf129ng-1 KRYNPQHRX

```

35

In addition, ORF129ng-1 (SEQ ID NO: 844) is homologous to an ABC transporter (SEQ ID NO: 1165) from *A.fulgidus*:

40

```

2650409(AE001090) glutamine ABC transporter, permease protein (glnP) [Archaeoglobus
fulgidus]Length = 224
Score = 132 bits (329), Expect = 2e-30
Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

Query: 65 VSLLYVT LFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRYGPIIAGSLAL 124
+S YV + RGTP L VQI+I +F P+ GI + E A G +AL
Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99

```

45

Query: 125 IANSGAYICEIFRAGIQSIDKGQMEAACSLGLTYPQAMRYVILPQALRRMLPPLASEFIT 184
 SGAYI EI RAGI+SI GQMEAA SLG+TY QAMRYVI PQA R +LP L +EFI
 Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAARSLGMTYLQAMRYVIFPQAFRNILPALGNEFIA 159

Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLEKR 242
 LLKDSSLLSVI++ EL V I P AL YL+MT L + +K+
 Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 845>] (SEQ ID NO: 845):

```

1  ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
101 GGCTGCCCCG GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151 CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCTT TACGTAAACA
201 cTACGTCCGC ACTTATTACy TGCTCCAAC TTTGCCCC GCAGgcTagT
251 TTGTGGACAG GCGCGCGGwa ATTACAAAC CTGCCCCGcYt CCGCGCCCCCT
301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
351 TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACCTCGA CTACCCCAAA
401 CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
451 TTTCTTGrTG AACGTGAACC CGrTATTTTT CATTACCGTT CCTGCGATTC
501 TGACCGCCGC CGTATTCGTA CTGTATCTTT TCrCGTTTAT ACCGATATTT
551 CGGCGGAATG CGTTTACAGA CGATCCGGA Tar

```

This corresponds to the amino acid sequence [<SEQ ID 846; ORF130>] (SEQ ID NO: 846; ORF130):

```

1  ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPAAPL
101 HLITLGGMMG GVMVWLTA LWHSGFTKLD YPKLCRIAVP ILFAAAVSRA
151 FLXNVNPXFF ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE*

```

Further work revealed the complete nucleotide sequence [<SEQ ID 847>] (SEQ ID NO: 847):

```

1  ATGCGGCCGT TTTTCGTCGG CGCGCGGGTG CTTGCCATAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101 TGGAACCTAT GCTGCCGGCG GCATACGGCG GTTTTTTGAC TGCGGCTTTG
151 TTGGA CTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCG CGACTTTGAT
201 GCGCGCATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
251 CTGCCTCGTT TTTCTGCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301 GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
351 GTTACTTGCC GCGTCACTG TTTTTCAGAC GGCATATGCC GTCAGCGCGC
401 ATTTGAACCT GTTGC GCGCG CAAGTGCATC TAAATATGGC GCGCGTGATG
451 TTCGTATCCG TGCGCGTCAG TATTCTTTTG GCGCGGAAG CCCTGAAAGA
501 ATGCCGTCTG AAAGACCCTG TTTTATTCC AAATATCGTT TATAAAAACA

```

5	551	TCGCCATTAC	TTTCTGCTC	TTGCACGCCG	CCGCCGAAC	TTGGCTGCC
	601	GCGCAAACCG	CCGGTTTAC	CGCGCTCGCC	GTCCGGTTCA	TCCTGCTCGC
	651	CAAGCTGCGT	GAGCTTCACC	ATCACGAACT	CTTACGTA	CACTACGTCC
	701	GCACTTATTA	CCTGCTCCAA	CTCTTTGCCG	CCGAGGCTA	TTTGTGGACA
	751	GGCGCGCGCA	AATTACAAAA	CCTGCCCGCC	TCCGCGCCCC	TGCACCTGAT
	801	TACCTCGGC	GGCATGATGG	GCGGCGTGAT	GATGGTGTGG	CTGACCGCCG
	851	GACTGTGGCA	CAGCGGCTTT	ACCAAAC	ACTACCCCA	ACTCTGCCG
	901	ATTGCCGTCC	CCATCTTTT	CGCCGCCGC	GTCTCGCGC	CTTCTTGAT
10	951	GAACGTGAAC	CCGATATTT	TCATTACCGT	TCCTCGGATT	CTGACCGCCG
	1001	CCGTATTCGT	ACTGTATCTT	TTCACGTTA	TACCGATATT	TCGGGCGAAT
	1051	GCGTTTACAG	ACGATCCGGA	ATAA		

This corresponds to the amino acid sequence [[<SEQ ID 848; ORF130-1>](#)] ([SEQ ID NO: 848; ORF130-1](#)):

15 1 MRPFFVGAAGV LAILGALVFF INPGAIVLHR QIFLELMPLA AYGGFLTAAL
 51 LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLLLFC
 101 ARLIWLDNRNT DNFALLMLLA AFTVFQTAYA VSGDNLNLLRA QVHLNMAAVM
 151 FVSVRVSIILL GAEALKECRL KDPVFIPNIV YKNIAITFLL LHAAAEWLWP
 201 AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYYLLQ LFAAAGYLWT
20 251 GAAKLQNLPA SAPLHLITLG GMMGGVMMVW LTAGLWHSGF TKLDYPKLCR
 301 IAVPILFAAA VSRAFLMNVN PIFFITVPAI LTAAEVFLYL FTFPIPIFRAN
 351 AFTDDPE*

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 (SEQ ID NO: 846) shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) (SEQ ID NO: 850) from strain A of *N. meningitidis*:

								10	20	30
orfl30.pep	LKECRLKDPVFIPNIVYKNIAITFLLHAA	:	:	:	:	:	:	:	:	:
orfl30a	LNLLRAQVHLNMAAVMFVSVRVISILLGAEALKECRLKDPVFIPNVVYKNIAITFLLHAA	:	:	:	:	:	:	:	:	:
	140	150	160	170	180	190				
		40	50	60	70	80	90			
orfl30.pep	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX	:	:	:	:	:	:	:	:	:
orfl30a	AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX	:	:	:	:	:	:	:	:	:
	200	210	220	230	240	250				
		100	110	120	130	140	150			
orfl30.pep	LQNLPASAPHLHLITLGMMGGVMWVLTAGLWHSGFTKLDPKLCRIAVPILFAAAVSRA	:	:	:	:	:	:	:	:	:
orfl30a	LQNLPASAPHLHLITLGMMGSVMWVLTAGLWHSGFTKLDPKLCRIAVPILFAAAVSRA	:	:	:	:	:	:	:	:	:
	260	270	280	290	300	310				
		160	170	180	190					
orfl30.pep	FLXNVNPXFFITVPAILTAADFVLYLFXXFIPIFRANAFTDDPEX	:	:	:	:	:	:	:	:	:
orfl30a	VLMNVNPIFFITVPAILTAADFVLYLLTFVPIFRANAFTDDPEX	:	:	:	:	:	:	:	:	:
	320	330	340	350						

The complete length ORF130a nucleotide sequence [<SEQ ID 849>] (SEQ ID NO: 849) is:

```

1  ATGCGGCCGT  TTTTCGTCGG  CGCGGCCGGT  CTTGCCATAC  TCGGTGCGCT
51  GGTGTTTTTC  ATCAACCCCG  GTGCCATCGT  CCTGCACCGC  CAAATTTTCT
5  101  TTGGAATTAT  GCTGCCGGCG  GCATACGGCG  GTTTTTTGAC  TGCGGCTTTG
151  TTGGAATTGA  CGGGTTTTTC  GGGTAACCTG  AAACCTGTCT  CGACTTTGAT
201  GGCGGCATT  TTGCTCGCCG  CATCCGCTAT  ACTGCCCTTT  TCGCCGCAAA
251  CTGCTCGT  TTTTCGTCGG  GCCTATTGGC  TGGTGTGCT  GCTGTTCTGC
301  GCCCGGCTGA  TTTGGCTAGA  CCGAAACACC  GACAACTTCG  CCCTGCTAAT
10  351  GTTACTTGCC  GCGTTCCTG  TTTTTCAGAC  GGCATATGCC  GTCAGCGGCG
401  ATTTGAACCT  GTTGCAGCG  CAAGTGCATC  TAAATATGGC  GCGGCTGATG
451  TTCGTATCCG  TGCGCGTCAG  TATTCTTTTG  GGCGCGGAAG  CCCTGAAAGA
501  ATGCCGCTG  AAAGACCCAG  TATTCATCCC  CAATGTCGTC  TATAAAAAACA
551  TCGCCATTAC  CTTCTGCTC  CTGCACGCCG  CCGCCGAAGT  TTGGCTGCCT
15  601  GCGCAAACCG  CCGGTTTTAC  CTCGCTCGCC  GTCGGCTTTA  TCCTGCTTGC
651  CAAGCTGCGT  GAGCTTACAC  ATCACGAACT  CCTGCGCAAA  CACTACGTCC
701  GCACTTATTA  CCTGTCCAA  CTCTTTGCCG  CCGCAGGCTA  TTTGTGGACA
751  GCGCGGCGCA  AATTACAAA  CCTGCCCGCC  TCCGCGCCCC  TGCACCTGAT
801  TACCCTCGGT  GGCATGATGG  GCAGCGTGAT  GATGGTGTGG  CTGACTGCCG
20  851  GACTGTGCA  CAGCGGCTTT  ACCAAGCTCG  ACTACCCGAA  ACTCTGCCGC
901  ATCGCCGTCC  CCATCCTNTT  CGCCGCCGCC  GTTTCGCGCG  CTGTTTTAAT
951  GAACGTAAAC  CCGATATTCT  TCATCACCGT  CCCCAGCAAT  CTGACCGCCG
1001  CCGTGTTCGT  GCTTTACCTG  CTGACATTCG  TACCGATCTT  TCGGGCGAAC
25  1051  GCGTTTACAG  ACGATCCGGA  ATAA

```

This encodes a protein having amino acid sequence [<SEQ ID 850>] (SEQ ID NO: 850):

```

1  MRPFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVL LLLFC
30  101  ARLIWLDRNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151  FVSVRSILL GAEALKECRL KDPVFIPNVV YKNIAITFLL LHAAELWLP
201  AQTAGFTSLA VGFILLAKLR ELHHHELLRK HYVRTYYLLQ LFAAGYLTW
251  GAAKLQNLPA SAPLHLITLG GMMGSVMVWV LTAGLWHS GF TKLDYPKLCR
301  IAVPILFAAA VSRVLMNVN PIFFITVPAI LTAAVFVLYL LTFVPIFRAN
35  351  AFTDDPE*

```

ORF130a (SEQ ID NO: 850) and ORF130-1 (SEQ ID NO: 848) show 98.3% identity in 357 aa overlap:

```

40  orf130a.pep  MRPFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL LDWTGFSGNL
orf130-1      MRPFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL LDWTGFSGNL

orf130a.pep  KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLDRNTDNFALLMLLA
orf130-1      KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLDRNTDNFALLMLLA

45  orf130a.pep  AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVV
orf130-1      AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVV

orf130a.pep  YKNIAITFLL LHAAELWLP AQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
orf130-1      YKNIAITFLL LHAAELWLP AQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQ

```


orf130a.pep LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSFGFTKLDYPKLCR
 orf130-1 LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSFGFTKLDYPKLCR
 5 orf130a.pep IAVPILFAAAVSRAVLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPE
 orf130-1 IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFVLYLFTFIPIFRANAFTDDPE

Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 (SEQ ID NO: 846) shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) (SEQ ID NO: 852) from *N.gonorrhoeae*:

10 orf130.pep LKECRLKDPVFIPNIVYKNIAITFLLHAA 30
 orf130ng LNLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNIVYKNIAIT-LLLHAA 201
 orf130.pep AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX 90
 15 orf130ng AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAX 261
 orf130.pep LQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSFGFTKLDYPKLCRIAVPILFAAAVSRA 150
 orf130ng LQNLPAAPLHLITLGGMTGGVMVWLTAGLWHSFGFTKLDYPKLCRIAVSILFASAVSRA 321
 orf130.pep FLXNVNPFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPE 193
 20 orf130ng VLMNVNPIFFITVPEILTAAVFMLYLLTFVPIFRANAFTDDPE 364

An ORF130ng nucleotide sequence [<SEQ ID 851>] (SEQ ID NO: 851) was predicted to encode a protein having amino acid sequence [<SEQ ID 852>] (SEQ ID NO: 852):

25 1 MNKFFTHPMR PFFVGA AVLA ILGALVFFHQ PRRYHPAPPN FLGTYAAGCI
 51 RRFDDYRFVG PDGFFRQ PET CRYFDGGVVA CCGCFIAVET ATCRIFRRRL
 101 LAGVA AVLRL ADLARRQHRT LRSVDVTA AF TVFQTAYAVS GDLNLLRAQV
 151 HLNMAAVMFV SVRVSVLLGT ETLKECRLKD PVFIPNVIYK NIAITLLHHA
 201 AAELWLPAQT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYLLQLFA
 30 251 AAGYLWTGAA KLQNLPAAP LHLITLGGMT GGVMVWLT GLWHSFGFTK
 301 DYPKLCRIAV SILFASAVSR AVLNVNPIF FITVPEILTA AVFMLYLLTF
 351 VPIFRANAFT DDPE*

Further work revealed the following gonococcal DNA sequence [<SEQ ID 853>] (SEQ ID NO:

35 853):

1 ATGCGCCCGT TTTTCGTCGG TCGGCAGTA CTGCCATAC TCGGTGCGTT
 51 GGTGTTTTT ATCAACCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT
 101 TGGAACCTAT GCTGCCGGCT GCATACGCG GTTTTTTGAC TACCGCTTG
 151 TTGGACCGGA CGGGTTTTTC AGGCAACCTG AAACCTGCCG CTACTTTGAT
 201 GCGGTGTTG TTGCTTGTG CCGCTGTTT ATTGCCGTTT TTACCGCAAC
 40 251 TTGCCGATT TTTTCGTCGC GCCTATTGGC TGGTGTGCT GCTGTCTGC
 301 GCCTGGCTGA TTTGGCTCGA CCGCAACACC GACAACTTCG CTCTGTTGAT
 351 GTTACTTGCC GCATTTACCG TTTTTCAGAC GGCCTATGCC GTCAGCGCGC

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 101

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 855>] (SEQ ID NO: 855):

```

1  ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
51  TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTTGGGATATT
151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGGA
351 CTGCTTGGA AAG..

```

This corresponds to the amino acid sequence [<SEQ ID 856; ORF131>] (SEQ ID NO: 856; ORF131):

```

1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51  GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSYFYRKIG KFEXCGLDWR
101 TRDGKPLIET FKQGGFDCLE K..

```

Further work revealed the complete nucleotide sequence [<SEQ ID 857>] (SEQ ID NO: 857):

```

1  ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
51  TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTTGGGATATT
151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGGA
351 CTGCTTGGA AAGCAGGGGT TGCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGGTAA

```

This corresponds to the amino acid sequence [<SEQ ID 858; ORF131-1>] (SEQ ID NO: 858; ORF131-1):

```

1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51  GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 (SEQ ID NO: 856) shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) (SEQ ID NO: 860) from strain A of *N. meningitidis*:

```

5      10      20      30      40      50      60
orf131.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
|
orf131a    MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLGWCCKPRKPAAIDFWDIGGESPPSLED
|
10      70      80      90      100     110     120
orf131.pep YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL
|
orf131a    YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
|
15      orf131.pep K
|
orf131a    KQGLRRNGLSERVRWX
|
130

```

The complete length ORF131a nucleotide sequence [<SEQ ID 859>] (SEQ ID NO: 859) is:

```

20      1  ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
51      51  TACGGTTGCA GGCTGCCGGT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT
101     101  CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
151     151  GGC GCGGAGA GTCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201     201  CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
25      251  ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
301     301  ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA
351     351  TTGTTTGAAA AAGCAGGGGT TCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401     401  GATGGTAA

```

30 This encodes a protein having amino acid sequence [<SEQ ID 860>] (SEQ ID NO: 860):

```

1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW*

```

35 ORF131a (SEQ ID NO: 860) and ORF131-1 (SEQ ID NO: 858) show 97.0% identity in 135 aa overlap:

```

40      orf131a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLGWCCKPRKPAAIDFWDIGGESPPSLED
|
orf131-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
|
40      orf131a.pep YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
|
orf131-1 YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCL
|
orf131a.pep KQGLRRNGLSERVRWX
|
45      orf131-1 KQGLRRNGLSERVRWX
|

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 (SEQ ID NO: 856) shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) (SEQ ID NO: 862) from *N.gonorrhoeae*:

```

5      orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  60
      orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECLSLSGWCKPRKPAAIDFWDIGGESPLSLED  60

      orf131.pep  YEIPLSDGNSSVRANEYESAQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE  120
      orf131ng    YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE  120

10     orf131.pep  K                                121
      orf131ng    KQGLRRNGLSERVRW  134

```

A complete length ORF131ng nucleotide sequence [<SEQ ID 861>] (SEQ ID NO: 861) was predicted to encode a protein having amino acid sequence [<SEQ ID 862>] (SEQ ID NO: 862):

```

1  MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence [<SEQ ID 863>] (SEQ ID NO: 863):

```

1  ATGGAAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTCATT
51 TACGGTTGCA GGCTGCCGGC TGGCGGGTG GTATGAGTGT TCGTCCTTGT
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
25 151 GCGGCGGAGA GtccgctGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT
251 ACTTTTATAG GAAATAGGG AAGTTTGAAG CCTGCGGGT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA
30 351 CTGTTTGAA AAGCAGGGT TCGGCGCAA CGGCCTGTCC GAGCGCGTCC
401 GATGGTAA

```

This corresponds to the amino acid sequence [<SEQ ID 864; ORF131ng-1>] (SEQ ID NO: 864; ORF131ng-1):

```

35 1  MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

ORF131ng-1 (SEQ ID NO: 864) and ORF131-1 (SEQ ID NO: 858) show 92.6% identity in 135 aa overlap:

```

40 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED

```

```

orf131-1      |||||:||||| |||:||||| ||||| ||||| |||||:||||| ||||| ||||| |||||
MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAIDFWDIGGESPPSLGD

orf131ng-1.pep YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf131-1      YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

orf131ng-1.pep KQGLRRNGLSERVRWX
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf131-1      KQGLRRNGLSERVRWX
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 865>] (SEQ ID
15 NO: 865)

	1	ATGAAACACA	TCCATATTAT	CGGTATCGGC	GGCAGCTTTA	TGGGCGGGCT
	51	TGCCGCCATT	GCCAAAGAAG	CGGGGTTTGA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	TATAGACGTG
20	151	TATGAAGGCT	TCGATGCCGC	TCAGTTGGAC	GAATTTAAAG	CCGACGTTTA
	201	CGTTATCGGC	AATGTCGCCA	AGCGCGGGAT	GGATGTGGTT	GAAGCGATTT
	251	TGAACCTCGG	CCTGCCtTAT	AtTtCcCGGC	CGCAATGGCT	GTCCGAAAAC
	301	GTGCTGCACC	ATCATTGGGT	ACTCGGTGTG	CGGGGACgC	ACGGCAAAC
	351	GACCACCGCC	TCCATGCTCG	CATGGGTCTT	GGAATATgCC	GGCCTCGCGC
25	401	CGGGCTTCCT	TATtGGCGGC	GTACC.GGAA	AATtCGGCG	TTTCCGCCCG
	451	CCTGCCGCAA	ACGCCGCGCC	AAGACCCGAA	CAGCCAATCG	CCGTTTTTcG
	501	TCATCGAAGC	CGACGAATAC	GACACCGCCT	TTtTCGACAA	ACGTTCATAA
	551	TtCGTGcATT	ACCGTCCGCG	TACCGCCGTG	TTGAACAATC	TGGAATTCTGA
	601	CCACGCCGAC	ATCTTTGCCG	ACTTGGGGCG	GATACAGACc	CAGTTCCACT
30	651	ACCTGCTGCG	TACCGTGCCG	TCTGAAGGCT	TAATCGTCTG	CAACGGACGG
	701	CAGCAAAGCC	TGCAAGATAC	TTTGGACAAA	GGCTGCTGGA	CGCCGGTGGA
	751	AAAATTCCGGC	ACGGAACACG	GCTGCA..		

This corresponds to the amino acid sequence [[<SEQ ID 866; ORF132>](#)] ([SEQ ID NO: 866; ORF132](#)):

35 1 MKHIHIIGIG GTFMGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
 51 YEGFDAAQLD EFKADVYVIG NVAKRGMDDV EAILNLGLPY ISGPQWLSN
 101 VLHHHWWLGV AGTHGKTTA SMLAWVLEYA GLAPFLIGG VXGKFRFRP
 151 PAANAAPRPE QPIAVFRHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
40 201 PRRHLCRLGR DTDVPVLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAG
 251 KIRHGTRLA..

Further work revealed the complete nucleotide sequence [<SEQ ID 867>] (SEQ ID NO: 867):

1 ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51 TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA

5
10
15
20
25

```

101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CCTGCCTTAT ATTTCCGGCC CGCAATGGCT GTCGGAAAAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
401 CGGGCTTCCT TATTGGCGGC GTACCGGAAA ATTTCCGGCT TTCCGCCCGC
451 CTGCCGCAAA CGCCGCGCCA AGACCGAAC AGCCAATCGC CGTTTTTCGT
501 CATTGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CGTTCTAAAT
551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTTCGAC
601 CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCCACTA
651 CCTCGTGCCT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
701 AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGTGGGAA
751 AAATTCCGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
801 CTCGTTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
851 ATTTGATGGG CAGGCAACAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
901 GCGCGTCATG TCGGTGTCGA TATTCAGACC GCCTGCGAAG CCTTGGGCGC
951 GTTTAAAAAC GTCAAACGCC GGATGGAAAT CAAAGGCACG GCAAACGGCA
1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
1051 ATTCAAGGTT TGCGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCTT
1101 CGAACCGCGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
1151 CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGCGGTG
1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
1351 GGAAAGCTGC TGGAAGCTTT GAGATAG

```

This corresponds to the amino acid sequence [<SEQ ID 868; ORF132-1>] (SEQ ID NO: 868; ORF132-1):

30
35
40

```

1 MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51 YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
201 HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE
251 KFGTEHGWQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
301 ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT
351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLEKA DQVFCYAGGV
401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH
451 GKLEALR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical o457 protein (SEQ ID NO: 1166) of *E.coli* (accession number U14003)

ORF132 (SEQ ID NO: 866) and o457 (SEQ ID NO: 1166) show 58% aa identity in 140 aa overlap:

45

```

Orf132: 4  IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLDEFK 63
           IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +
o457: 3  IHIILGICGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61

```

Orf132: 64 ADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTASML 123
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121

Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143
 W+LE G PGF+IGGV G
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF132 (SEQ ID NO: 866) shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) (SEQ ID NO: 870) from strain A of *N. meningitidis*:

10		10	20	30	40	50	60
	orf132.pep	MKHIHIIGIGGTFMGG	LAAIAKEAGFEVSG	CDAKMYPPMSTQ	LEALGIDVYEGF	DAAQLD	
	orf132a	MKHIHIIGIGGTFMGG	LAAIAKEAGFEXSG	CDAKMYPPMSTQ	LEALGIGVYEGF	DTAQLD	
		10	20	30	40	50	60
15		70	80	90	100	110	120
	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA					
	orf132a	EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA					
		70	80	90	100	110	120
20		130	140	150	160		
	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFR---	RFRPPAANAAPRPEQPI-----	AVFR			
	orf132a	SMLAWVLEYAGLAPGFXIGGVPENFSVSARL-	PQTPRQDPNSQSPFFVIEADEYDTAFFD				
		130	140	150	160	170	
25		170	180	190	200	210	220
	orf132.pep	HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHL	CRLGRDTPVPLPRAYRAVXRL				
	orf132a	KRSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLV	RTVPSEGLIVCNGRQQSLQD				
30		180	190	200	210	220	230

The complete length ORF132a nucleotide sequence [<SEQ ID 869>] (SEQ ID NO: 869) is:

1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGTGGGAT
51	TGCCGCCATT	GCCAAAGAAG	CAGGGTTTGA	ANTCAGCGGT	TGCGATGCGA
101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTG
151	TATGAAGGCT	TCGACACCGC	GCAGTTGGAC	GAATTTAAAG	CCGACGTTTA
201	CGTTATCGGC	AATGTCGCCA	AGCGCGGGAT	GGATGTGGTT	GAAGCGATTT
251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAC
301	NTGCTGCACC	ATCATTGGNN	ACTCGGCGTG	GCGGNGACGC	ACGGCAAAAC
351	GACCACCGCG	TCTATGCTCG	CGTGGGTTT	GGAATATGCC	GGACTCGCAC
401	CGGGCTTCNT	TATCGGCGGC	GTACCGGAAA	ACTTCAGCGT	TTCCGCCCGC
451	CTGCCGCAAA	CGCCGCGCCA	AGACCCGAAC	AGCCAATCGC	CGTTTTTCGT
501	CATTGAAGCC	GACGAATACG	ACACCGCGTT	TTTCGACAAA	CGCTCCAAAT
551	TCGTGCATTA	CCGTCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
601	CACGCCGACA	TCTTCGCCGA	TTTGGGCGCG	ATACAGACCC	AGTTCACCA
651	CCTCGTGCGT	ACCGTGCCGT	CTGAAGGCCT	CATCGTCTGC	AACGGACGGC
701	AGCAAAGCCT	GCAAGAACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTCGAA
751	AAATTCGGCA	CGGAACACGG	CTGGCAGGCC	GGCGAAGCCA	ATGCCGATGG
801	CTCGTTCGAC	GTGTTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTCGCTTGGA

5
10
15
20
25
30
35
40
45

```

851  GTTTGATGGG  CGGACACAAC  CGCATGAACG  CGCTCGCNGT  CATCGCCGCC
901  GCGCGTCATG  CCGGAGTNGA  CATTGAGACG  GCCTGCGAAG  CCTTGAGCAC
951  GTTTAAAAAC  GTCAAACGCC  GCATGGAAAT  CAAAGGCACG  GCAAACGGTA
1001 TCACCGTTTA  CGACGACTTC  GCCCACCATC  CGACCGCTAT  CGAAACCACG
1051 ATTCAAGGTT  TGCGCCAGCG  CGTCGGCGGC  GCGCGCATCC  TCGCCGTCCT
1101 CGAACCGCGT  TCCAATACGA  TGAAGCTGGG  TACGATGAAA  GCCGCCCTGC
1151 CCGCAAGCCT  CAAAGAAGCC  GACCAAGTGT  TCTGNTACGC  CGGCGGCGCG
1201 GACTGGGACG  TTGCCGAAGC  CCTCGCGCCT  TTGGGCGGCA  GGCTGCACGT
1251 CGGCAAAGAC  TTCGATGCCT  TCGTTGCCGA  AATCGTGAAA  AACGCCGAAG
1301 CAGGCGACCA  TATTTTGGTG  ATGAGCAACG  GCGGTTTCGG  CGGAATACAC
1351 ACCAACTGC   TGGACGCTTT  GAGATAG

```

This encodes a protein having amino acid sequence [<SEQ ID 870>] (SEQ ID NO: 870):

15
20
25
30
35
40
45

```

1  MKHIHIIGIG  GTFMGGIAAI  AKEAGFEXSG  CDAKMYPPMS  TQLEALGIGV
51  YEGFDTAQLD  EFKADVYVIG  NVAKRGMDEV  EAILNRGLPY  ISGPQWLAEN
101 XLHHHWXLGV  AXTHGKTTTA  SMLAWVLEYA  GLAPGFXIGG  VPENFVSAR
151 LPQTPRQDPN  SQSPFFVIEA  DEYDTAFDDK  RSKFVHYRPR  TAVLNNLEFD
201 HADIFADLGA  IQTQFHHLVR  TVPSEGLIVC  NGRQQSLQDT  LDKGCWTPVE
251 KPGTEHWQA   GEANADGSFD  VLLDGKKAGH  VAWSLMGGHN  RMNALAVIAA
301 ARHAGVDIQT  ACEALSTFKN  VKRMEIKGT   ANGITVYDDF  AHHPTAIETT
351 IQGLRQRVGG  ARILAVLEPR  SNTMKLGTMK  AALPASLKEA  DQVFXYAGGA
401 DWDVAEALAP  LGGRLHVGKD  FDAFVAEIVK  NAEAGDHILV  MSNGGFGGIH
451 TKLLDALR*

```

25 ORF132a (SEQ ID NO: 870) and ORF132-1 (SEQ ID NO: 868) show 93.9% identity in 458 aa overlap:

30
35
40
45

```

orf132a.pep  MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD
|||||:|||||
orf132-1     MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD

orf132a.pep  EFKADVYVIGNVAKRGMDEVVEAILNRGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA
|||||:|||||
orf132-1     EFKADVYVIGNVAKRGMDEVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA

orf132a.pep  SMLAWVLEYAGLAPGFXIGGVPEFVSARLPQTPRQDPNSQSPFFVIEADEYDTAFDDK
|||||:|||||
orf132-1     SMLAWVLEYAGLAPGFLIGGVPEFVSARLPQTPRQDPNSQSPFFVIEADEYDTAFDDK

orf132a.pep  RSKFVHYRPRRTAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQSLQDT
|||||:|||||
orf132-1     RSKFVHYRPRRTAVLNNLEFDHADIFADLGAIQTQFHYLVRTVPSEGLIVCNGRQQSLQDT

orf132a.pep  LDKGCWTPVEKFGTEHWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA
|||||:|||||
orf132-1     LDKGCWTPVEKFGTEHWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA

orf132a.pep  ARHAGVDIQTACEALSTFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
|||:|||||
orf132-1     ARHVGVDIQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

orf132a.pep  ARILAVLEPRSNTMKLGTMKAAALPASLKEADQVFXYAGGADWDVAEALAPLGGRLHVGKD
|||||:|||||
orf132-1     ARILAVLEPRSNTMKLGTMKSAALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD

```

```

orf132a.pep    FDAFVAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX
               |||||:|||||:|||||:|||||:|||||:
orf132-1       FDAFVAEIVKNAEVDHILVMSNGGFGGIHGKLLLEALRX

```

Homology with a predicted ORF from *N.gonorrhoeae*

- 5 ORF132 (SEQ ID NO: 866) shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) (SEQ ID NO: 872) from *N. gonorrhoeae*:

```

orf132.pep    MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD    60
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf132ng      MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQLE    60

10 orf132.pep    EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA    120
               ||:|:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf132ng      EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA    120

15 orf132.pep    SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ    180
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf132ng      SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ    180

orf132.pep    TFXIRALPSAYRRVEQSGIRPRRHLRRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY    240
               |: |||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf132ng      TLQIRALSPAYRRVEQSGIRPRRHLRRLGRDTPVPPRAHRTIRRPHRLQRTAAKPARY    240

20 orf132.pep    FGQRLLDAGGKIRHGTRLA    259
               |||||:|||||:|||||:|||||:|||||:
orf132ng      FGQRLLDAGGKIRHRTLADW    261

```

An ORF132ng nucleotide sequence [<SEQ ID 871>] (SEQ ID NO: 871) was predicted to encode a protein having amino acid sequence [<SEQ ID 872>] (SEQ ID NO: 872):

```

25      1 MKHIHIIGIG GTFMGGIAAI AKEAGFKVSG CDAKMYPPMS TQLEALGIGV
      51 HEGFDAAQLE EFQADIYVIG NVARRGMDVV EAILNRGLPY ISGPQWLAEN
     101 VLHHHWLVGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPGKFRFRFP
     151 PTANAASRPE QQIAVFRHRS RRIRHRLFRQ TLQIRALSPA YRRVEQSGIR
     201 PRRHLRRLGR DTPVPPPPRA HRTIRRPHRL QRTAAKPARY FGQRLLDAGG
30     251 KIRHRTLAD W*

```

Further work revealed the following gonococcal DNA sequence [<SEQ ID 873>] (SEQ ID NO: 873):

```

35      1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGAT
      51 TGCCGCCATT GCCAAGAAG CCGGGTTCAA AGTCAGCGGT TGCGACGCGA
     101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTA
     151 CACGAAGGCT TCGATGCCGC GCAGTTGGAA GAATTTCAAG CCGATATTTA
     201 CGTCATCGGC AATGTCGCCA GGCGCGGGAT GGATGTGGTC GAGGCGATT
40     251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAac
     301 GTGctgcacc atcaTTGGgt ACTCGGCGTG GcaggaCGC ACGGcaaAac
     351 gaccaCcGcg tCCATGCTCG CCTGGGTCTT GGAATATGCC GGACTCGCGC
     401 CGGGCTTCCT CATCGGCGGt gtaccggaAA ATTTCCGGCT TTCCGCCCGC
     451 CTACCGCAAA CGCCGCGTCA AGACCCGAAC AGCAAATCGC CGTTTTTCGT

```



```

orf132ng-1.pep  ARHAGVDVQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
|||:||||:|||||||||||||||||||||||||||||||||||||||||
orf132-1        ARHVGVDIQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

orf132ng-1.pep  ARILAVLEPRSNMTKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRLRVGKD
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf132-1        ARILAVLEPRSNMTKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD

orf132ng-1.pep  FDTFVAEIVKNARTGDHILVMSNGGFGGIIHTKLLDALRX
||:|||||:|||||:|||||:|||||:|||||:|||||
orf132-1        FDAFVAEIVKNAEVDGHILVMSNGGFGGIIHGKLEALRX

```

In addition, ORF132ng-1 (SEQ ID NO: 874) is homologous to a hypothetical *E.coli* protein (SEQ ID NO: 1166):

pir||S56459 hypothetical protein o457 - Escherichia coli)gi|537075 (U14003)
ORF_o457 [Escherichia coli])gi|1790680 (AE000494) hypothetical 48.5 kD protein in
fbp-pmba intergenic region [Escherichia coli] Length = 457
Score = 474 bits (1207), Expect = e-133
Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

Query: 22 KEAGFKVSGCDAKMYPPPMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
 ++ G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE
 Sbjct: 21 RQLGHEVTGSDANVYPPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

Query: 82 AILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141
A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV
Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCGYKPGFVIGGV 139

Query: 142 PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRTAVLNNLEFDH 201
P NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH
Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

Query: 202 ADIFADLGAIQQTQFHHLVRTVPSEGLIVCNGQQQSLQDITLDKGCWTPVEKFGTGHGWQIG 261
ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ
Sbjct: 191 ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWQAK 250

Query: 262 EVNADGS-FDVLDDGKKAGHVAWDLMGGHNRMNALAVIAAARHAGVDVQTACEALGAFKN 320
 ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N
 Sbjct: 251 KLTTDASEWEVLLDGEKVGVEKWSLVGEHNMHNGLMAIAAARHVGVPADAANALGSFIN 310

Query: 321 VKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMTMKLGTM 379
+RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNMTMK+G
Sbjct: 311 ARRRLELRGEANGVTVYDDFAHHPTAILATLAALRGKVGGTARIIVAVLEPRSNMTMKMGIC 370

Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRLRVGKDFDTFVAEIVKNARTGDHI 438
K L SL AD+VF W VAE D DT +VK A+ GDHI
Sbjct: 371 KDDLAPSLGRADEVFLQLQPAHIPWQVAEVAEACVQPAHWSGVDVTLADMVVKTAQPGDHI 430

Query: 439 LVMSNGGFGGIIHTKLILDAL 457
 LVMSNGGFGGGIH KLLD L
 Sbjct: 431 LVMSNGGFGGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF132-1 (SEQ ID NO: 868) (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These experiments confirm that ORF132 (SEQ ID NO: 866) is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 875>] (SEQ ID NO: 875)

```
1  .CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
51  CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
101 CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
151 ATTAGTGCGG ACTTCGGCGA TTATTTTCATG CCGTTCGCCA GCTATTCGCG
201 CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTTCC CAAATCGGCG
251 ACTCCGGCGT TCACACCGCC TTAAAACCAG AGCGCGCAAA CACTTGGCAA
301 TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAAAACAAG ATGATACATT
351 AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
401 ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTCC GAGCTGGGTC
451 AGCAGCACCG GGCTTGCCCTA CACCATCCAA CATCGCrATT TCawAGACAA
501 AGTGCATCAA nnnnnnnnnnn nnnnnnnnnnn nnnnnTACGAT TATGGGCGTT
551 TTTTCACCAA CCTTTCTTAC GCCTATCAA AAAGCACGCA ACCGACCAAC
601 TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
651 CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
701 GACGTTTGGA AGTCGGTACG CGCTGGTTGG GCAACAAACT GACTTTGGGC
751 GGC GCGATGC GCTATTTCTG CAAGAGCATC CGCGCGACGG CTGAAGAACG
801 CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTT CGGCAACTGG
851 GCAAGCGTTC CATCAAACAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
901 TTWgATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTTCCGCGC
951 CGAAGTCAAA AATCTGTTCT ACAGGCGTTA TATCGATCCG CTCGATGCGG
1001 GCAATGATGC GGCAAC .GAG CGTTATTACA GCTCGTTCGA CCCGAAAGAC
1051 AAGGACrRAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
1151 CCTTTTTgAT GACGATGAGC TACAAGTTTT AA
```

This corresponds to the amino acid sequence [<SEQ ID 876; ORF133>] (SEQ ID NO: 876; ORF133):

```
1  .PGYYGSDDEF KRAFGENSPT XKKHCNRSCG IYEPVLKKYG KKRANNHSVS
51  ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
101 FGFXTYKKGL LKQDDTLGLK LVGYRSRIDN YIHNVYKWW DLNGDIPSWV
151 SSTGLAYTIQ HRFXDKVHQ XXXXXXXXXYD YGRFFTNLSY AYQKSTQPTN
201 FSDASEPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTLG
251 GAMRYFGKSI RATAEERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
301 XDFNAAYEPK KNLI FRAEVK NLFDRRYIDP LDAGNDAAAXE RYSSSFDPKD
```

351 KDXDVTGNAD KTLNKGKYG TSKSVLTNFA RGRFTLMTMS YKF*

Further work revealed the further partial DNA sequence [SEQ ID 877] (SEQ ID NO: 877):

```
5      1  GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
      51  ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
     101  AGGATATATT CAAATCCAGC GAAACCTCG ACAACATCGT ACGCAGCATC
     151  CCCGGTGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
     201  TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA
     251  TCACGCAGAC CTTTATTTCG ACTTCTACCG ATGCGGGCAG GGCAGGCGGT
    301  TCATCTCAAT TCGGTGCATC TGTCGACAGC AATTTTATTG CCGGACTGGA
    351  TGTCGTCAAA GGCAGCTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG
    401  GTTCGCGGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGGCAAT
    451  AATACCTACG GCCTGCTGCT AAAAGGTCTG ACCGGCACCA ATTCAACCAA
    501  AGGTAATGCG ATGGCGGCGA TAGGTGCGCG CAAATGGCTG GAAAGCGGAG
    551  CATCTGTGCG TGTGCTTTAC GGGCACAGCA GGCACAGCGT GCGCAAAAT
    601  TACCGCGTGG GCGGCGGCGG GCAGCACATC GGAAATTTTG GCGCGGAATA
    651  TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT
    701  TCAATCCGA CAGCGAAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG
    751  AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAAGTAC AaAAATACAT
    801  CGAAGAGCAT GACAAAAGCT GCGGGAAAAA CCTg. CaCCG CAATACGACA
    851  TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG
    901  TTTAAATTGG AATACGACGG CGTATTCAAT AAATACACGG CGCAATTTCTG
    951  CGATTTAAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTATC
   1001  AGTTCAATTA CGGTTTGTCT TTGAACCCGT ATACCAACCT CAATCTGACC
   1051  GCAGCCTACA ATTCGGGCAG GCAGAAATAT CCGAAAGGGT CGAAGTTTAC
   1101  AGGCTGGGGG CTTTTAAAGG ATTTTGAAAC CTACAACAAC GCGAAAAATCC
   1151  TCGACCTCAA CAACACCGCC ACCTTCCGGC TGCCCCGCGA AACCAGAGTTG
   1201  CAAACCACTT TGGGCTTCAA TTATTCCAC AACGAATACG GCAAAAACCG
   1251  CTTTCCTGAA GAATTGGGGC TGTTTTCGA CGGTCCTGAT CAGGACAACG
   1301  GGCTTTATTC CTATTGGGGG CGGTTTAAGG GCGATAAAGG GCTGCTGCCC
   1351  CAAAAATCAA CCATTGTCCA ACCGGCCGGC AGCCAATATT TCAACACGTT
   1401  CTACTTCGAT GCCGCGCTCA AAAAAAGCAT TTACCGCTTA AACTACAGCA
   1451  CCAATACCGT CGGCTACCGT TTCGGCGGGC AATATACGGG CTATTACGGC
   1501  TCGGATGACG AATTTAAGCG GGCATTGCGA GAAAACTCGC CGACATACAA
   1551  GAAACATTGC AACCGGAGCT GCGGGATTTA TGAACCCGTA TTGAAAAAAT
   1601  ACGGCAAAAA GCGCGCCAAC AACCATTGCG TCAGCATTAG TGCGGACTTC
   1651  GCGGATTATT TCATGCCGTT CGCCAGCTAT TCGCGCACAC ACCGTATGCC
   1701  CAACATCCAA GAAATGTATT TTTCCCAAAT CGGCGACTCC GGCCTTCACA
   1751  CCGCCTTAAA ACCAGAGCGC GCAAACACTT GGCAATTTGG CTTCAATACC
   1801  TATAAAAAAG GATTGTTAAA ACAAGATGAT ACATTAGGAT TAAAAGTGGT
   1851  CGGCTACCGC AGCCGCATCG ACAACTACAT CCACAACGTT TACGGGAAAT
   1901  GGTGGGATTT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT
   1951  GCCTACACCA TCCAACATCG CAATTTCAAA GACAAAGTGC ACAAACACGG
   2001  TTTTGAGTTG GAGCTGAATT ACGATTATGG GCGTTTTTTC ACCAACCTTT
   2051  CTTACGCCTA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC
   2101  GAATCGCCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG
   2151  GTTGAGCAGG GTTTCGCCCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG
   2201  GTACGCGCTG GTTGGGCAAC AAAGTACTT TGGGCGGCGC GATGCGCTAT
   2251  TTCCGCAAGA GTTCCGCGC GACGGCTGAA GAACGCTATA TCGACGGCAC
   2301  CAACGGGGGA AATACCAGCA ATTTCCGGCA ACTGGGCAAG CGTTCCATCA
   2351  AACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTGA TTTTACGCC
   2401  GCTTACGAGC CGAAGAAAAA CCTTATTTTC CGCGCGAAG TCAAAAATCT
   2451  GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA
   2501  CGCAGCGTTA TTACAGCTCG TTCGACCCGA AAGACAAGGA CGAAGACGTA
   2551  ACGTGTAATG CTGATAAAAC GTTGTGCAAC GGCAATACG GCGGCACAAG
   2601  CAAAAGCGTA TTGACCAATT TTGCACGCGG ACGCACCTTT TTGATGACGA
   2651  TGAGCTACAA GTTTTAA
```

This corresponds to the amino acid sequence [<SEQ ID 878; ORF133-1>] (SEQ ID NO: 878; ORF133-1):

```

      1  EAQIQVLEDV  HVKAKRVPKD  KKVFTDARAV  STRQDIFKSS  ENLDNIVRSI
      51  PGAFQTQDKS  SGIVSLNIRG  DSGFGRVNTM  VDGITQTFYS  TSTDAGRAGG
5      101  SSQFGASVDS  NFIAGLDVVK  GSFSGSAGIN  SLAGSANLRT  LGVDDVVQGN
      151  NTYGLLLKGL  TGTNSTKGNA  MAAIGARKWL  ESGASVGVLY  GHSRRSVAQN
      201  YRVGGGGQHI  GNFGAEYLER  RKQRYFVQEG  ALKFNSDSGK  WERDLQRQQW
      251  KYKPYKNYNN  QELQKYIEEH  DKSRENLXP  QYDITPIDPS  SLKQQSAGNL
      301  FKLEYDGVFN  KYTAQFRDLN  TKIGSRKIIN  RNYQFNYGLS  LNPYTNLNL
      351  AAYNSGRQKY  PKGSKFTGWG  LLKDFETYNN  AKILDNLNTA  TFRLPRETEL
10     401  QTTLGFNYFH  NEYGKNRFPF  ELGLFFDGPD  QDNGLYSYLG  RFKGDKGLLP
      451  QKSTIVQPAG  SQYFNTFYFD  AALKKDIYRL  NYSTNTVGYR  FGGEYTGYYG
      501  SDDEFKRAFG  ENSPTYKKHC  NRSCGIYEPV  LKKYGKKRAN  NHSVSISADF
      551  GDYFMPFASY  SRTHRMPNIQ  EMYFSQIGDS  GVHTALKPER  ANTWQGFNT
15     601  YKKGLLKQDD  TLGLKLVGYR  SRIDNYIHN  YGKWWDLNGD  IPSWVSSTGL
      651  AYTQHRNFK  DKVHKHGFEL  ELNYDYGRFF  TNLSYAYQKS  TQPTNFSNAS
      701  ESPNNASKED  QLKQGYGLSR  VSALPRDYGR  LEVGTRWLGN  KLTLLGAMRY
      751  FGKSIRATAE  ERYIDGTNGG  NTSNFRQLGK  RSIKQTETLA  RQPLIFDFYA
      801  AYEPPKNLIF  RAEVKNLFDR  RYIDPLDAGN  DAATQRYYSS  FDPKDKDEDV
20     851  TCNADKTLN  GKYGGTSKSV  LTNFARGRTF  LMTMSYKF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801) (SEQ ID NO: 1167)

25 ORF133 (SEQ ID NO: 876) and HI121 (SEQ ID NO: 1167) show 57% aa identity in 363aa overlap:

```

Orf133: 31  IYEPVLKKYGKKRANNHNSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
          I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA
HI121: 563  INEPILHKSGHKKAFNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFFSQVSNAGVNTA 622

30  Orf133: 91  LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150
          LKPE+++T+Q GF TYKKGL QDD LG+KLVGYRS I NYIHNVYG WW +P+W
HI121: 623  LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYIHNVYGWW--RDGMPTWA 680

Orf133: 151  SSTGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSNASESPNN 210
          S G YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN
35  HI121: 681  ESNQFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSAYQRTNQPTNYADASPRPNN 740

Orf133: 211  ASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTLLGAMRYFGKSIRATAEERYID 270
          AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLLG A RY+GKS RAT EE YI+
HI121: 741  ASQEDILKQGYGLSRVSMLPKDYGRLELGTRWFDQKLTLLGLAARYYGKSKRATIEEEYIN 800

Orf133: 271  GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRIYIDP 330
          G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP
40  HI121: 801  GSR-FKKNLTLRRENYAVKKTEDIKKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDP 859

```

```

Orf133: 331 LDAGNDAAXERYSSFPDKDKXDVTCTNADKTLCTNGKYGGTSSKSVLTNFAFGRTFLMTMS 390
          LDAGNDAA +RYYSS      + + C D + C      GG+ K+VL NFARGRT++++++
HI121:  860 LDAGNDAASQRYSSL-----NNSIECAQDSSAC----GGSDKTVLYNFARGRTYILSLN 910

Orf133: 391 YKF 393
          YKF
HI121:  911 YKF 913

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 (SEQ ID NO: 876) shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) (SEQ ID NO: 880) from strain A of *N. meningitidis*:

[illegible]

orf133.pep KFX
 |||
orf133a KFX
 870

5

A partial ORF133a nucleotide sequence [<SEQ ID 879>] (SEQ ID NO: 879) is:

10

15

20

25

30

35

40

45

50

55

```
1  AAAGACAAAA AAGTGTTTAC CGATGCGCGT GCCGTATCGA CCCGTCAGGA
51 TATATTCAAA TCCANCGAAA ACCTCGACAA CATCGTACGC ANCATCCCCG
101 GTGCGTTTAC ACANCAANAT AAAAGCTCGG GCNTTGTGTC TTTGAATATT
151 CGCNGCGACA GCGGGTTCGG GCGGGTCAAT ACNATGGTNG ACGGCATCAC
201 NCANACCTTT TATTCGACTT CTACCGATGC GGGCAGGGCA GCGGGTTCAT
251 CTCAAATCGG TGCATCTGTC GACAGCAATT TTATNGCCGG ACTGGATGTC
301 GTCAAAGGCA GCTTCAGCGG CTCGGCAGGC ATCAACAGCC TTGCCGGTTC
351 GGCGAATCTG CGGACTTTAN GCGTGGATGA TGTCGTTTCA GGCAATANTA
401 CNTACGCGCT GCTGCTAAAA GGTCTGACCG GCACCAATTC AACCAAAGGT
451 AATGCGATGG CCGCGATAGG TGCGCGCAAA TGGCTGGAAA GCGGAGCATC
501 TGTCGGTGTG CTTTACGGGC ACAGCAGGCG CAGCGTGGCG CAAAATTACC
551 GCGTGGGCGG CCGCGGGCAG CACATCGGAA ATTTTGGCGC GGAATATCTG
601 GAACGACGCA AGCAACGATA TTTTGAGCAA GAAGGCGGGT TGAAATTCAA
651 TTCCAACAGC GGAATATGGG AGCGGGATTT CCAAAAGTCG TACTGGAAAA
701 CCAAGTGGTA TCAAAAATAC GATGCCCCC AAGAACTGCA AAAATACATC
751 GAAGTCATG ATAAAAGCTG GCGGGAAGAC CTGGCGCCGC AATACGACAT
801 CACCCCATC GATCCGTCCA GCCTGAAGCN GCAGTCGGCA GGCAACCTGT
851 TTAAATTGGA ATACGACGGC GTATTCAATA AATACACGGC GCAATTTTCG
901 GATTTAAACA CCAAATCGG CAGCCGCAAA ATCATCAACC GCAATTATCA
951 ATTCAATTAC GGTTTGTCTT TGAACCCGTA TACCAACCTC AATCTGACCG
1001 CAGCCTACAA TTCGGGCAGG CAGAAATATC CGAAAGGGTC GAAGTTTACA
1051 GGCTGGGGGC TTTTNAAGA TTTTGAAACC TACAACAACG CAAAATCCT
1101 CGACCTCANC AACACCTCCA CCTTCCGGCT GCCCGTGAA ACCGAGTTGC
1151 AAACCACTTT GGGCTTCAAT TATTTCCACA ACGAATACGG CAAAAACCGC
1201 TTTCTGAAG AATTGGGGCT GTTTTTCGAC GGTCCGGATC ANGACAACGG
1251 GCTTTATTC TATTTGGGGC GGTTTAAGGG CGATAAAGGG CTGCTGCCCC
1301 AAAAAACAAC CATTTGCCAA CCGGCCGGCA GCCAATATTT CAACACGTT
1351 TACTTCGATG CCGCGCTCAA AAAAGACATT TACCGCTTAA ACTACAGCAC
1401 CAATACCGTC GGCTACCGTT TCGGCGGCNA ATATACGGC TATTACNGCT
1451 CGGATGACGA ATTTAAGCGG GCATTCCGAG AAAACTCGCC GACATACANG
1501 AAACATTGCA ACCAGAGCTG CGGAATTTAT GAACCCGTAT TGAAAAAATA
1551 CGGCAAAAAG CGCGCCAACA ACCATTCCGT CAGCATTAGT GCGGACTTCG
1601 GCGATTATTT CATGCCGTTT GCCAGCTATT CGCGCACACA CCGTATGCCC
1651 AACATCCAAG AAATGTATTT TTCCCAAATC GGCGACTCCG CCGTTCACAC
1701 CGCCTTAAAA CCAGAGCGCG CAACACTTG GCAATTTGGC TTCAATACCT
1751 ATAAAAAAGG ATTGTTAAAA CAAGATGATA TATTAGGATT AAAACTGGTC
1801 GGCTACCGCA GCCGCATCGA CNACTACATC CACAACGTTT ACGGGAAATG
1851 GTGGGATTTG AACGGGAATA TTCCGAGCTG GGTGAGCAGC ACCGGGCTTG
1901 CCTACACCAT CCAACACCGC AATTTCAAAG ACAAGTGCA CAAACACGGT
1951 TTTGAGTTGG AGCTGAATTA CGATTATNGG CGTTTTTTCA CCAACCTTTC
2001 TTACGCCTAT CAAAAAAGCA CGCAACCGAC CAACTTCAGC GATGCGAGCG
2051 AATCGCCCAA CAATGCGTCC AAAGAAGACC AACTCAAACA AGGTTATGGG
2101 TTGAGCAGGG TTTCCGCCCT GCCGCGAGAT TACGAGCTT TGGAAGTCGG
2151 TACGCGCTGG TTGGGCAACA AACTGACTTT GGGCGGCGCG ATGCGCTATT
2201 TCGGCAAGAG CATCCGCGCG ACGGCTGAAG AACGCTATAT CGACGNCACC
2251 AATGGGNAN NTACCAGCAA TTTCCGCAA CTGGGCAAGC GTTCCATCAN
2301 ACAAACCGAA ACCCTTGCCC GCCAGCCTTT GATTTTTGAT TTNTACGCCG
2351 CTTACGAGCC GAAGAAAAAN CTTATTTTCC GCGCCGAAGT CAAAAATCTG
2401 TTCGACAGGC GTTATATCGA TCCGCTCGAT GCGGGCAATG ATGCGGCAAC
2451 GCAGCGTTAT TACAGTTCGT TCGACCCGAA AGACAAGGAC GAAGAAGTAA
2501 CGTGTAATGA TGATAACACG TTATGCAACG GCAAATACGG CGGCACAAGC
2551 AAAAGCGTAT TGACCAATTT TGCACGCGGA CNCACCTTTT TGATAACGAT
2601 GAGCTACAAG TTTTAA
```

This encodes a protein having (partial) amino acid sequence [<SEQ ID 880>] (SEQ ID NO: 880):

```

      1  KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI
      5  51  RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
      10  101 VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
      15  151 NAMAAGARK  WLESGASGV  LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
      20  201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWKYQKY DAPQELQKYI
      25  251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR
      30  301 DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT
      35  351 GWGLXKDFET YNNAKILDLX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR
      40  401 FPEELGLFFD GPDKDNGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
      45  451 YFDAALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX
      50  501 KHCNQSCGIY EPVLKKYGKK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP
      55  551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLV
      60  601 GYRSRIDXYI HNVYGKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
      65  651 FELELNIDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG
      70  701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
      75  751 NGXXTSNFRQ LGRKSIXQTE TLRQPLIFD XYAAYEPKKX LIFRAEVKNL
      80  801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNT LCNGKYGGTS
      85  851 KSVLTNFARG XTFLITMSYK F*
  
```

ORF133a (SEQ ID NO: 880) and ORF133-1 (SEQ ID NO: 878) show 94.3% identity in 871 aa overlap:

```

      25  orf133a.pep      10      20      30      40
      orf133-1      EAQIQVLEDVHVKA RVPKDKKVFTDARAVSTRQDIFKSS ENLDNIVRSIPGAFTQQDKS
      30  orf133a.pep      50      60      70      80      90      100
      orf133-1      SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGGSSQFGASVDSNFXAGLDVVK
      35  orf133a.pep      110     120     130     140     150     160
      orf133-1      GSFSGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAGARKWL
      40  orf133a.pep      170     180     190     200     210     220
      orf133-1      ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
      45  orf133a.pep      230     240     250     260     270     280
      orf133-1      WERDFQKSYWTKWKYQKYDAPQELQKYIEGHDKSWRENLAPQYDITPIDPSSLKXQSAGN
      50  orf133a.pep      290     300     310     320     330     340
      orf133-1      LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
  
```

	orf133-1	 LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYSLSNPNYTNLNLTAAYNSGRQK	300	310	320	330	340	350
5	orf133a.pep	350 360 370 380 390 400 YPKGSKFTGWGLXKDFETYNNAKILDLXNTSTFRLPRETELQTTLGFNYPHNEYGKNRFP						
	orf133-1	 YPKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELQTTLGFNYPHNEYGKNRFP	360	370	380	390	400	410
10	orf133a.pep	410 420 430 440 450 460 EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR						
	orf133-1	 EELGLFFDGPQDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR	420	430	440	450	460	470
15	orf133a.pep	470 480 490 500 510 520 LNYSTNTVGYRFGGXVTGYXSDDEFKRAFGENSPTYXKHCNQSCGIYEPVLKKYGKKRA						
	orf133-1	 LNYSTNTVGYRFGGEYTGYYGSDDEFKRAFGENSPTYXKHCNRSCGIYEPVLKKYGKKRA	480	490	500	510	520	530
20	orf133a.pep	530 540 550 560 570 580 NNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN						
	orf133-1	 NNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN	540	550	560	570	580	590
25	orf133a.pep	590 600 610 620 630 640 TYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVSSTGLAYTIQHRNF						
	orf133-1	 TYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVSSTGLAYTIQHRNF	600	610	620	630	640	650
30	orf133a.pep	650 660 670 680 690 700 KDKVHKHGFELNLDYXRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS						
	orf133-1	 KDKVHKHGFELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS	660	670	680	690	700	710
35	orf133a.pep	710 720 730 740 750 760 RVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDXTNGXXTSNFRQLG						
	orf133-1	 RVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRQLG	720	730	740	750	760	770
40	orf133a.pep	770 780 790 800 810 820 KRSIXQTETLARQPLIFDXAAYEPKXLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS						
	orf133-1	 KRSIXQTETLARQPLIFDYAAYEPKKNLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS	780	790	800	810	820	830
45	orf133a.pep	830 840 850 860 870 SFDPKDKDEEVTCTNDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX						
	orf133-1	 SFDPKDKDEDVTCNADKTLCTNGKYGGTSKSVLTNFARGRTFLMTMSYKFX	840	850	860	870	880	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF133 (SEQ ID NO: 876) shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) (SEQ ID NO: 882) from *N. gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPTXKKHCNRSCGI	31
5	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
	orf133.pep	YEPVLKKYGGKRRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGGKRRANNHVSISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
10	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
	orf133.pep	STGLAYTIQHRXFXDKVHQXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	211
	orf133ng	STGLAYTIRHRNFKDKVHKHGFLELNYDYGRFFTNLSYAYQKSTQPTNFSDAESPNNNA	740
15	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRLGNKLTGGAMRYFGKSI RATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRLGNKLTGGAMRYFGKSI RATAEERYIDG	800
	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL	331
20	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAAYEPKKNLIFRAEVKNLFDRRYIDPL	860
	orf133.pep	DAGNDAAXERYYSSFDPKDKDXDVT CNADKTL CNGKYGGTSKSVLTNFARGRTFLMTMSY	391
	orf133ng	DAGNDAATQRYYSSFDPKDKDEDTV CNADKTL CNGKYGGTSKSVLTNFARGRTFLMTMSY	920
25	orf133.pep	KF 393	
	orf133ng	KF 922	

The complete length ORF133ng nucleotide sequence [[<SEQ ID 881>](#)] ([SEQ ID NO: 881](#)) is predicted to encode a protein having amino acid sequence [[<SEQ ID 882>](#)] ([SEQ ID NO: 882](#)):

30	1	MRSSFRLKPI	CFYLMGVMLY	HHSYAEDAGR	AGSEAIQIVL	EDVHVKAKRVR
	51	PKDKKVFTDA	RAVSTRQDVF	KSGENLDNIV	RSIPGAFTQQ	DKSSGIVSLN
	101	IRGDSGFGRV	NTMVDGITQT	FYSTSTDAGR	AGSSSQFGAS	VDSNFIAGLD
	151	VVKGSFSGSA	GINSLAGSAN	LRTLGVDDVV	QGNNTYGLLL	KGLTGTNSTK
	201	GNAMAAIGAR	KWLESGASVG	VLYGHSRRGV	AQNYRVGGGG	QHIGNFGEEY
35	251	LERRKQYFV	QEGGLKFNAG	SGKWERDLQR	QYWKTKWYKK	YEDPQELQKY
	301	IEEHDKSWRE	NLAPQYDITP	IDPSGLKQOS	AGNLLNLEYD	GVFNKYTAQF
	351	RDLNTRIGSR	KIINRNYQFN	YGLSLNPYTN	LNLTAAYNSG	RQKYPKGAKF
	401	TGWGLLKDFE	TYNNAKILDL	NNTATFRLPR	ETELQTTLGF	NYFHNEYGKN
	451	RFPEELGLFF	DGPDQDNGLY	SYLGRFKGDK	GLLPQKSTIV	QPAGSQYFNT
40	501	FYFDAALKKD	IYRLNYSTNA	INYRFGGEYT	GYGSENEFK	RAFGENSPAY
	551	KEHCDPSCGL	YEPVLKKYGK	KRANNHSVSI	SADFGDYFMP	FAGYSRTHRM
	601	PNIQEMYFSQ	IGDSGVHTAL	KPERANTWQF	GFNTYKKGLL	KQDDILGLKL
	651	VGYRSRIDNY	IHNVYGKWWD	LNGDIPSWVG	STGLAYTIRH	RNFKDKVHKH
	701	GFELELNIDY	GRFFTNLSYA	YQKSTQPTNF	SDASESPNNA	SKEDQLKQGY
45	751	GLSRVSALPR	DYGRLEVGT	WLGKLTLLGG	AMRYFGKISR	ATAEERYIDG
	801	TNGGNTSNVR	OLGKRSIKOT	ETLAROLPIF	DFYAAYEPKK	NLIFRAEVKN

851 LFDRRYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK TLCNGKYGGT
 901 SKSVLTNFAR GRTFLMTMSY KF*

A variant was also identified, being encoded by the gonococcal DNA sequence [<SEQ ID 883>]

5 (SEQ ID NO: 883):

1 ATGAGATCTT CTTTCCGGTT GAAGCCGATT TGTTTTTATC TTATGGGTGT
 51 TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGGCAGCG
 101 AGGCGCAGAT ACAGGTTTTG GAAGATGTGC ACGTCAAGGC GAAGCGCGTA
 151 CCGAAAGACA AAAAAGTGTT TACCGATGCG CGTGCCGTAT CGACCCGTca
 201 gGATGTGTTC AAATCCGGCG AAAACCTCGA CAACATCGTA CGCAGCATAC
 251 CCGGTGCGTT TACACAGCAA GATAAAAGCT CGGGCATTGT GTCTTTGAAT
 301 ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT
 351 CACGCAGACC TTTTATTCTGA CTTCTACCGA TGCGGGCAGG GCAGGCGGTT
 401 CATCTCAATT CGGTGCATCT GTCGACAGCA ATTTTATTGC CGGACTGGAT
 451 GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG
 501 TTCGCGCAAT CTGCGGACTT TAGGCGTGGA TGACGTGCTT CAGGGCAATA
 551 ATACCTACGG CCTGCTGCTA AAAGGTCTGA CCGGCACCAA TTCAACCAAA
 601 GGTAATGCGA TGGCGGCGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC
 651 GTCTGTCCGT GTGCTTTACG GGCACAGCAG GCGCGGCGTG GCGCAAAATT
 701 ACCGCGTGGG CCGCGGCGGG CAGCACATCG GAAATTTTGG TGAAGAATAT
 751 CTGGAACGGC GCAAACAGCA ATATTTTGTA CAAGAGGGTG GTTTGAAATT
 801 CAATGCCGGC AGCGGAAAAAT GGGAACGGGA TTTGCAAAGG CAATACTGGA
 851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCCAAGAAGT GCAAAAATAC
 901 ATCGAAGAGC ATGATAAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA
 951 CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG CGAGGCAATC
 1001 TGTTTAAATT GGAATACGAC GGCCTATTCA ATAAATACAC GGCGAATTT
 1051 CCGGATTAA ACACCAGAAT CGGCAGCCGC AAAATCATCA ACCGCAATTA
 1101 TCAATTCAAT TACGGTTTGT CTTTGAACCC GTATACCAAC CTCAATCTGA
 1151 CCGCAGCCTA CAATTCGGGC AGGCAGAAAT ATCCGAAAGG GCGGAAGTTT
 1201 ACAGGCTGGG GGCTTTTAAA AGATTTTGAA ACCTACAACA ACGCGAAAAT
 1251 CCTCGACCTC AACACACCG CCACCTTCCG GCTGCCCCGC GAAACCGAGT
 1301 TGCAAACCCAC TTTGGGCTTC AATTATTTCC ACAACGAATA CGGCAAAAAC
 1351 CGCTTTCCTG AAGAATTGGG GCTGTTTTTC GACGGTCCTG ATCAGGACAA
 1401 CGGGCTTTAT TCCTATTTGG GGCGGTTTAA GGGCGATAAA GGCTGTTGC
 1451 CTCAAAAATC AACCATTGTC CAACCGGCCG GCAGCCAATA TTTCAACACG
 1501 TTCTACTTCG ATGCCGCGCT CAAAAAAGAC ATTTACCGCT TAAACTACAG
 1551 CACCAATGCA ATCAACTACC GTTTCGGCGG CGAATATACG GGCTATTACG
 1601 GCTCGGAAAA CGAATTTAAG CGGGCATTTC GAGAAAACCT GCCGGCATAC
 1651 AAGGAACATT GCGACCCGAG CTGCGGGCTT TATGAACCCG TATTGAAAAA
 1701 ATACGGCAAA AAGCGCGCCA ACAACCATTG GGTGAGCATT AGTGCGGACT
 1751 TCGGCGATTA TTTCATGCGG TTCGCGGCTT ATTCGCGCAC ACACCGTATG
 1801 CCCAACATCC AAGAAATGTA TTTTTCCTCA ATCGGCGACT CCGGCGTTCA
 1851 CACCGCCTTA AAACCAGAGC GCGCAAACAC TTGGCAATTT GGCTTCAATA
 1901 CCTATAAAAA AGGATTGTGA AAACAAGATG ATATATTAGG ATGAAACTG
 1951 GTCGGCTACC GCAGCCGCAT TGACAACTAC ATCCACAACG TTACGGGAA
 2001 ATGGTGGGAT TTGAACGGGG ATATTCCGAG CTGGGTCGGC AGCACCGGGC
 2051 TTGCCTACAC CATCCGACAC CGCAATTTCA AAGACAAAGT GCACAAACAC
 2101 GGTTTTGAGC TGGAGCTGAA TTACGATTAT GGGCGTTTTT TCACCAACCT
 2151 TTCTTACGCC TATCAAAAAA GCACGCAACC GACCAATTTT AGCGATGCGA
 2201 GCGAATCGCC CAACAATGCC tccaaAGAAG ACCAACTCAA ACAAGGTTAT
 2251 GGGCTGAGCA GGGTTTCCGC CCTGCCGCGA GATTACGGAC GTTTGGAAGT
 2301 CGGTACGCGC TGGTTGGGCA ACAAACCTGAC TTTGGGCGGC GCGAtgCGCT
 2351 ATTTCCGGCA GAGCATCCGC GCGACGGCTG AAGAACGCTA TATCGACGGC
 2401 ACCAACGGGG GAAATACCGA CAATGTCCGG CAACTGGGCA AGCGTTCCAT
 2451 CAAACAAACC GAAACCCCTTG CCCGACAGCC TTTGATTTTT GATTTTTTACG
 2501 CCGCTTACGA GCCGAAGAAA AACCTTATTT TCCGCGCCGA AGTCAAAAAC
 2551 CTGTTTCGACA GCGGTATATAT CGATCCGCTC GATGCGGGCA ATGATGCGGC
 2601 AACGCAGCGT TATTACAGCT CGTTCGACCC GAAAGACAAG GACGAAGACG

2651 TAACGTGTAA TGCTGATAAA ACGTTGTGCA ACGGCAAATA CGGCGGCACA
 2701 AGCAAAAGCG TATTGACCAA TTTCGCACGC GGACGCACCT TCTTGATGAC
 2751 GATGAGCTAC AAGTTTAA

5 This corresponds to the amino acid sequence [<SEQ ID 884; ORF133ng-1>] (SEQ ID NO: 884;
ORF133ng-1):

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAIQVL EDVHVKAERV
 51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
 10 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGGSSQFGAS VDSNFIAGLD
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
 251 LERRKQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQOS AGNLFKLEYD GVFNKYTAQF
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPTYN LNLTAAYNSG RQKYPKGAKF
 15 401 TGWGLLKDFE TYNNAKILD L NNTATFRLPR ETELQTTLGF NYFHNEYGKN
 451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
 501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYYGSENEFK RAFGENSPAY
 551 KEHCDPSCGL YEPVLKKYGK KRANNHSVSI SADFGDYFMP FAGYSRTHRM
 601 PNIQEMYFSQ IGDGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
 20 651 VGYRSRIDNY IHNVYGKWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH
 701 GFELELNVDY GRFFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
 751 GLSRVSALPR DYGRLEVGT R WLGNKLTGG AMRYFGKSIR ATAEERYIDG
 801 TNGGNTSNVR QLGRKSIQT ETLARQPLIF DFYAAEPPK NLIFRAEVKN
 851 LFDRRYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK TLCNGKYGGT
 25 901 SKSVLTNFAR GRTFLMTMSY KF*

ORF133ng-1 (SEQ ID NO: 884) and ORF133-1 (SEQ ID NO: 878) show 96.2% identity in 889 aa overlap:

30 orf133ng-1.pep 10 20 30 40 50 60
 orf133-1
 35 orf133ng-1.pep 70 80 90 100 110 120
 orf133-1
 40 orf133ng-1.pep 130 140 150 160 170 180
 orf133-1
 45 orf133ng-1.pep 190 200 210 220 230 240
 orf133-1
 50 orf133ng-1.pep 250 260 270 280 290 300

10

20

25

30

35

40

45

790 800 810 820 830 840
 orf133ng-1.pep YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
 |||||
 orf133-1 YFGKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
 750 760 770 780 790 800

-624-

```

      850      860      870      880      890      900
orf133ng-1.pep FRAEVKNLFDRRYIDPLDAGNDAATQRYYSSFDPKDKDEDVTCNADKTLNKGKYGSTKS
      |||
orf133-1      FRAEVKNLFDRRYIDPLDAGNDAATQRYYSSFDPKDKDEDVTCNADKTLNKGKYGSTKS
5      810      820      830      840      850      860

      910      920
orf133ng-1.pep VLTNFARGRTFLMTMSYKFX
      |||
orf133-1      VLTNFARGRTFLMTMSYKFX
10     870      880

```

In addition, ORF133ng-1 (SEQ ID NO: 884) is homologous to a TonB-dependent receptor (SEQ ID NO: 1167) in *H.influenzae*:

```

15  sp|P45114|YC17_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR
    )gi|1075372|pir||G64110 transferrin binding protein 1 precursor (tbpl) homolog -
    Haemophilus influenzae (strain Rd KW20) )gi|1574147 (U32801) transferrin binding
    protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913
    Score = 930 bits (2377), Expect = 0.0
    Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

20  Query: 38 QVLEDVHVKA KRVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97
    + L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V
    Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKGSGVV 88

    Query: 98 SLNIRGDSGFVRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFS 157
    S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS
25  Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGGSSQFGAIDPNFIAGVDVKNKSNFS 148

    Query: 158 GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217
    G++GIN+LAGSAN RTLGV+DV+ M RKWL++G
    Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLNDNGG 208

    Query: 218 SVGVLYGHSSRRGVAQNYRVGGGQHIGNFGEEYLERRKQQYFVQEGGLKFNAGSGKWERD 277
    VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D
30  Sbjct: 209 YGVVYGYGYSQREVSQDYRI-GGGERLASLGQDILAKEEAYF-RNAGYILNP-EGQWTPD 265

    Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303
    L +++W +Y KK +D ++LQK IEE
    Sbjct: 266 LSKKHWSCKNPKDYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKGNDEGIEE 325

35  Query: 304 HDKSWRENLA PQYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363
    DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI
    Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFYGDHQNLAGAQLRTLNDKIGSRKIE 384

    Query: 364 NRNYQFNYSLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT 423
    NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+
40  Sbjct: 385 NRNYQVNYNFNNNSYLDLNLMAAHNIGKTIYPKGGFFAGQVADKLITKNVANIVDINNS 444

    Query: 424 ATFRLPRETELQTTLGFNYPFHNEYGKNRFPPEELGLFFDGPDPQDNGLYSY--LGRFKGDKG 481
    TF LP+E +L+TTLGFNYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K
    Sbjct: 445 HTFLLPKEIDLKTTLGFNYPFTNEYSKNRFPPEELSLFYNDASHDQGLYSHSKRGRYSGTKS 504

    Query: 482 LLPQKSTIVQ PAGSQYFNTFYFDAALKKDIYRLNYSNAINYRFGGEYTGYYGSENEFKR 541
    LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY
45  Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDTALS KGIYHLNYSVNFTHYAFNGEYVGY----- 555

```


Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKYKGRANNHVSISADFGDYFMPFAGYSRTHRMP 601
 EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP
 Sbjct: 556 ---ENTAGQQ-----INEPILHKSGHKAFNHSATLSAELSDYFMPFFTYSRTHRMP 604

Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDILGLKLVGYRSRIDNYI 661
 NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKGL QDD+LG+KLVGYRS I NYI
 Sbjct: 605 NIQEMFFSQVSNAGVNTALKPEQSDTYQLGFNTYKGLFTQDDVLGVKLVGYRSFIKNYI 664

Query: 662 HNVYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNYDYGRFFTNLSYAY 721
 HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY
 Sbjct: 665 HNVYGVWW--RDGMPTWAESNGFKYITIAHQNYKPIVKKSGVELEINYDMGRFFANVSYAY 722

Query: 722 QKSTQPTNFSDASESPNNASKEDQKQGYGLSRVSALPRDYGRLEVGTWLGKLTGLGA 781
 Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTGL A
 Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLPKDYGRLELGTRWFDQKLTGLA 782

Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAEPKKN 841
 RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+
 Sbjct: 783 ARYYGKSKRATIEEEYINGSR-FKKNTLRRENYAVKKTEDIKKQPIILDHVSYPEIKD 841

Query: 842 LIFRAEVKNLFDRLYIDPLDAGNDAATQRYYSSFDPKDKDEDVTCNADKTLNKGKYGGS 901
 LI +AEV+NL D+RY+DPLDAGNDAA+QRYYSS + + C D + C GG+
 Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYYSSL-----NNSIECAQDSSAC---GGSD 892

Query: 902 KSVLTNFARGRTFLMTMSYKF 922
 K+VL NFARGRT+++++YKF
 Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 104

The following partial DNA sequence was identified in *N.meningitidis* [<SEQ ID 885>] (SEQ ID NO: 885)

30 1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
 51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
 101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
 151 GGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
 201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
 35 251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
 301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT
 351 CCGCGAATGG GTTGGCGCCA CACTGAGCCA AAAAGCCGAA AACATCAAA
 401 CCGCCGCCAT CAACGCAAA ATCAGCACCG GCAATACCGG CCTTGGCTG
 451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

This corresponds to the amino acid sequence [<SEQ ID 886; ORF112>] (SEQ ID NO: 886; ORF112):

1 MNLISRYIIR QMAVMVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
 51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAIN GK ISTGNTGLWL
 151 KEKNSVINVR EMLPDH...

Further work revealed further partial nucleotide sequence [<SEQ ID 887>] (SEQ ID NO: 887):

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
 51 TTACGCGCTC CTTCCTTTCC TCGCTTTGTA CAGCTTTTTC GAAATCCTGT
 101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
 151 gGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
 201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCGGCA
 251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
 301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
 351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
 401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
 451 AAAGAAAAAA ACAGCrTkat CAATGTGCGC GAAATGTTGC CCGACCATAC
 501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTGGCAG
 551 AGGCAGTGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGTTGGCAG
 601 TTGAAAAACA TCCGCCGAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC
 651 TATTGCGGCT GAAGAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG
 701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
 751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
 801 CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC
 851 TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC
 901 TTAATACTCT TCGGCGGCAT CTGTsTCGGA TTGCTGTTCC ACCTTGCCGG
 951 ACGGCTCTTT GGGTTTACCA GCCAACTCGG...

This corresponds to the amino acid sequence [<SEQ ID 888; ORF112-1>] (SEQ ID NO: 888; ORF112-1):

1 MNLISRYIIR QMAVMVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
 51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAIN GK ISTGNTGLWL
 151 KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
 201 LKNIRRTLGL EDKVEVSIAA EENWPISVKR NLMDVLLVKP DQMSVGELTT
 251 YIRHLQNSQ NTRIIAIAWW RKLVPAAAW VMALVAFAT PQTTRHGNMG
 301 LKLFGGICXG LLEHLAGRLF GFTSQL...

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 (SEQ ID NO: 886) shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) (SEQ ID NO: 890) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf112.pep	MNLISRYIIRQMAVMVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
orf112a	MNLISRYIIRQMAVMVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR					
	10	20	30	40	50	60

		70	80	90	100	110	120
	orf112.pep	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLILSQFGFIFAIATVALGEW					
		:		:			
5	orf112a	AYELMPLAVLIGGLVVSXSQLAAGSELXVIKASGMSTKKLLILSQFGFIFAIATVALGEW					
		70	80	90	100	110	120
		130	140	150	160		
	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH					
					:		
10	orf112a	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN					
		130	140	150	160	170	180
	orf112a	ELAEAVEADSAVLNSDGSWLKNIRRTLGEDKVEVSIAAEEXWPISVKRNLMDEVLLVKP					
		190	200	210	220	230	240

The ORF112a nucleotide sequence [<SEQ ID 889>] (SEQ ID NO: 889) is:

15	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCCG	GCCTACGAAC	TGATGCCCTT
	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCGCGCA
20	251	CGCAACTGAN	CGTCATCAAA	GCCAGCGGCA	TGAGACACAA	AAAGCTGCTG
	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTTT	GCTATTGCCA	CCGTCGCGCT
	351	CGGCGAATGG	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
25	501	CCTGCTGGGC	ATTAATAATCT	GGGCCGCAAA	CGATAAAAAC	GAACTGGCAG
	551	AGGCAGTGGA	AGCCGATTCC	GCCGTTTGA	ACAGCGACGG	CAGTTGGCAG
	601	TTGAAAAACA	TCCGCCGCAG	CACGCTTGGC	GAAGACAAAG	TCGAGGTCTC
	651	TATTGCGGCT	GAAGAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
	701	ACGTATTGCT	CGTCAAACCC	GACCAAAATGT	CCGTCGGCGA	ACTGACCACC
30	751	TACATCCGCG	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	CTTACGCCAT
	801	CGCATGGTGG	CGCAAATTGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCGTCGCCCT	TGCCTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
	901	TTAAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGTGTTTCC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCCTCG
35	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence [<SEQ ID 890>] (SEQ ID NO: 890):

40 1 MNLISRYIIR QMAVMAYVAL LAFLALYSFF EILYETGNLG KGSYGIWEMX
51 GYTALKMXAR AYELMPLAVL IGGLVSXSQ AAGSELXVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSIINV R EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
201 LKNIRRSTLG EDKVEVSIAA EEXWPISVKR NLMDVLLVKP DQMSVGELTT
251 YIRHLQXXSQ NTRIYAIAWW RKLVPAAAW VMALVAF AFT PQTRHGNMG
45 301 LKXFGGICLG LLFHLAGRLF XFTSQLYGIP PFLXGALPTI AFALLAVWLI
351 RKOEKR*

ORF112a (SEQ ID NO: 890) and ORF112-1 (SEQ ID NO: 888) show 96.3% identity in 326 aa overlap:

50 orf112a.pep MNLISRYII RQMAVMAYVALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR

[illegible]

Homology with a predicted ORF from *N.gonorrhoeae*

ORF112 (SEQ ID NO: 886) shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) (SEQ ID NO: 892) from *N. gonorrhoeae*:

20	orf112.pep	MNLISRYIIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
	orf112ng	MNLISRYIIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
	orf112.pep	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW	120
25	orf112ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLILSQFGFIFAIAAVALGEW	120
	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH	166
	orf112ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN	180

30 The complete length ORF112ng nucleotide sequence [SEQ ID 891] (SEQ ID NO: 891) is:

	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGC	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCTTCC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGCTG
35	151	GGCTACACCG	CCCTCAAAAT	GCCCGCCCGC	GCCTACGAAC	TCATGCCCCCT
	201	CGCCGTCCTC	ATCGGCGGAC	TGGCCTCTCT	CAGCCAGCTT	GCCGCGCGTA
	251	GCGAAGTGGC	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
	301	TTGATTCTGT	CTCAGTTCGG	TTTTATTTTT	GCTATTGCCG	CCGTGCGGCT
	351	CGGCGAATGG	GTTGCGCCCA	CGCTGAGCCA	AAAAGCCGAA	AACATCAAag
	401	cCGCCGCCat	taacggCAAA	ATCAGCAccg	gcAATACCGG	CCTTTggcTG
40	451	AAAGAAAAAa	ccAGCATTAT	CAATGTGcGc	GGAATGTTGC	CCGACCATAC
	501	GCTTTTGGGC	ATCAAAATTT	GGGCGCGCAA	CGATAAAAAA	GAATTGGCAG
	551	AGGCAGTGGA	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	CAGCTGGCAG
	601	TTGAAAAAa	TCCGCCGACG	CATCATGGGT	ACAGACAAAA	TCGAAACATC
	651	cqCCGCCGCC	GAGAAACTT	qGCCGATTGC	CGTCAGACGC	AACCTGATGG

5
 701 ACGTATTGCT CGTCAAGCCC GACCAAATGT CCGTCGGCGA GCTGACCACC
 751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCAAA TCTACGCCAT
 801 CGCATGGTGG CGTAAACTCG TTTACCCCGT CGCCGCATGG GTCATGGCGC
 851 TCGTTGCCTT CGCCTTTACG CCGCAAACCA CGCGCCACGG CAATATGGGC
 901 TTAAAACTCT TCGGCGGCAT CTGTCTCGGA TTGCTGTTCC ACCTTGCCGG
 951 CAGGCTCTTC GGGTTTACCA GCCAACTCTA CGGCACCCCA CCCTTCCTCG
 1001 CCGGCGCACT GCCTACCATA GCCTTCGCCT TGCTCGCTGT TTGGCTGATA
 1051 CGCAAACAGG AAAACGTTG A

10 This encodes a protein having amino acid sequence [SEQ ID 892] (SEQ ID NO: 892):

15
 1 MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
 51 GYTALKMPAR AYELMPLAVL IGGLASLSQL AAGSELAVIK ASGMSTKKLL
 101 LILSQGFIF AIAAVALGEW VAPTLSQKAE NIKAAINGK ISTGNTGLWL
 151 KEKTSIINVR GMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
 201 LKNIRRSIMG TDKIETSAAA EETWPIAVRR NLMDVLLVKP DQMSVGELTT
 251 YIRHLQNN SQ NTQIYAI AWW RKLVPVAAW VMALVAF AFT PQTRHGNMG
 301 LKLFGGICLG LLFHLA GRLF GFTSQLYGTP PFLAGALPTI AFALLAVWLI
 351 RKQEK R*

20 ORF112ng (SEQ ID NO: 892) and ORF112-1 (SEQ ID NO: 888) show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
	orf112ng	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
25	orf112-1	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
		10	20	30	40	50	60
	orf112ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLILSQGFIFAI A AVALGEW					
30	orf112-1	AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLILSQGFIFAIATVALGEW					
		70	80	90	100	110	120
	orf112ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLG IKIWARNDKN					
35	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLG IKIWARNDKN					
		130	140	150	160	170	180
	orf112ng	ELAEAVEADS AVLNSDGSWQLKNIRRSIMGTDKIETSAAA EETWPIAVRRNLMDVLLVKP					
40	orf112-1	ELAEAVEADS AVLNSDGSWQLKNIRRSTLGEDKVEVSIAAENWPISVKRNLMDVLLVKP					
		190	200	210	220	230	240
	orf112ng	DQMSVGELTTYIRHLQNN SQNTQIYAI AWW RKLVPVAAWVMALVAF AFT PQTRHGNMG					
45	orf112-1	DQMSVGELTTYIRHLQNN SQNTRIYAI AWW RKLVPAAAWVMALVAF AFT PQTRHGNMG					
		250	260	270	280	290	300
	orf112ng	LKLFGGICLGLL FHLA GRLFGFTSQLYGTPPFLAGALPTI AFALLAVWLIRKQEKRX					
50	orf112-1	LKLFGGICXGLL FHLA GRLFGFTSQL					

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 105

5

Table III lists several *Neisseria* strains which were used to assess the conservation of the sequence of ORF 4 (SEQ ID NO: 216) among different strains.

TABLE III – List of Neisseria Strains Used for Gene Variability Study of ORF 4 (SEQ ID NO: 216)

ORF4 gene variability: List of used Neisseria strains		
IdentificationStrains number		Source / reference
	Group B	
zv01_4	NG6/88	R. Moxon / Seiler et al., 1996
zv02_4	BZ198	R. Moxon / Seiler et al., 1996
zv03_4ass	NG3/88	R. Moxon / Seiler et al., 1996
zv04_4	297-0	R. Moxon / Seiler et al., 1996
zv05_4	1000	R. Moxon / Seiler et al., 1996
zv06_4	BZ147	R. Moxon / Seiler et al., 1996
zv07_4	BZ169	R. Moxon / Seiler et al., 1996
zv08_4	528	R. Moxon / Seiler et al., 1996
zv09_4	NGP165	R. Moxon / Seiler et al., 1996
zv10_4	BZ133	R. Moxon / Seiler et al., 1996
zv11_4	NGE31	R. Moxon / Seiler et al., 1996
zv12_4ass	NGF26	R. Moxon / Seiler et al., 1996
zv13_4	NGE28	R. Moxon / Seiler et al., 1996
zv15_4	SWZ107	R. Moxon / Seiler et al., 1996
zv16_4	NGH15	R. Moxon / Seiler et al., 1996
zv17_4	NGH36	R. Moxon / Seiler et al., 1996
zv18_4	BZ232	R. Moxon / Seiler et al., 1996
zv19_4	BZ83	R. Moxon / Seiler et al., 1996
zv20_4	44/76	R. Moxon / Seiler et al., 1996
zv21_4	MC58	R. Moxon
zv96_4	2996	Our collection
	Group A	
zv22_4	205900	R. Moxon
z2491_4	Z2491	R. Moxon / Maiden et al., 1998

	Group C	
zv24_4	90/18311	R. Moxon
zv25_4	93/4286	R. Moxon
	Others	
zv26_4ass	A22 (group W)	R. Moxon / Maiden et al., 1998
zv27_4	E26 (group X)	R. Moxon / Maiden et al., 1998
zv28_4	860800 (group Y)	R. Moxon / Maiden et al., 1998
zv29_4	E32 (group Z)	R. Moxon / Maiden et al., 1998
	Gonococcus	
zv32_4Ng F62	R. Moxon / Maiden et al., 1998	
zv33_4	Ng SN4	R. Moxon
fa1090_4	FA1090	R. Moxon

References:

Seiler A. *et al.*, Mol. Microbiol., 1996, 19(4):841-856.
Maiden et al., Proc. Natl. Acad. Sci. USA, 1998, 95:3140-3145.

The amino acid sequences for each listed strain are as follows:

- 5 >FA1090_4 [<SEQ ID 893>] (SEQ ID NO: 893)
MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVK
EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEAF
QVPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTAS
KADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYPAAWNEGAAK*
- 10 >Z2491_4 [<SEQ ID 894>] (SEQ ID NO: 894)
MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*
- 15 >ZV01_4 [<SEQ ID 895>] (SEQ ID NO: 895)
MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
20 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*
- >ZV02_4 [<SEQ ID 896>] (SEQ ID NO: 896)
MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
25 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAK*
- >ZV03_4ASS [<SEQ ID 897>] (SEQ ID NO: 897)
MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ

VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

5 >ZV04_4 [<SEQ ID 898>] (SEQ ID NO: 898)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

10 >ZV05_4 [<SEQ ID 899>] (SEQ ID NO: 899)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
15 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV06_4 [<SEQ ID 900>] (SEQ ID NO: 900)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
20 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTAHKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV07_4 [<SEQ ID 901>] (SEQ ID NO: 901)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
25 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV08_4 [<SEQ ID 902>] (SEQ ID NO: 1107)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
30 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV09_4 [<SEQ ID 902>] (SEQ ID NO: 902)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
35 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

40 >ZV10_4 [<SEQ ID 903>] (SEQ ID NO: 903)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
45 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV11_4 [<SEQ ID 904>] (SEQ ID NO: 904)
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE
QIQVELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
50 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV12_4ASS [<SEQ ID 905>] (SEQ ID NO: 905)
MKTFFKTLSSAAALALILAACGGQKDRAPAASASAAENGAKEILFGTTVGDLGDMVKE
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
55 ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV13_4 [<SEQ ID 906>] (SEQ ID NO: 906)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV15_4 [<SEQ ID 907>] (SEQ ID NO: 907)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAGNENGAAK*

>ZV16_4 [<SEQ ID 908>] (SEQ ID NO: 908)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV17_4 [<SEQ ID 909>] (SEQ ID NO: 909)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV18_4 [<SEQ ID 910>] (SEQ ID NO: 910)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV19_4 [<SEQ ID 911>] (SEQ ID NO: 911)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV20_4 [<SEQ ID 912>] (SEQ ID NO: 912)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV21_4 [<SEQ ID 913>] (SEQ ID NO: 913)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV22_4 [<SEQ ID 914>] (SEQ ID NO: 914)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDLVKE
QIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLKSLSEVVDGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

>ZV24_4ASS [<SEQ ID 915>] (SEQ ID NO: 915)

MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVLVEFTDDVRPNLALAEGLDIIIVQHKPYLDDFKKEQNLDITEVFQ

VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

5 >ZV25_4 [<SEQ ID 916>] (SEQ ID NO: 916)
MKTFFKTLSSAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
QIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

10 >ZV26_4 [<SEQ ID 917>] (SEQ ID NO: 917)
MKTFFKTLSSAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
15 AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

20 >ZV27_4 [<SEQ ID 918>] (SEQ ID NO: 918)
MKTFFKTLSSAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
QIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

25 >ZV28_4 [<SEQ ID 919>] (SEQ ID NO: 919)
MKTFFKTLSSAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

30 >ZV29_4 [<SEQ ID 920>] (SEQ ID NO: 920)
MKTFFKTLSSAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
QIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

35 >ZV32_4 [<SEQ ID 921>] (SEQ ID NO: 921)
MKTFFKTLSSAALALILAACGGQKDSAPAASAAAPSADNGAEKKEIVFGTTVGDFGDMVK
EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEAF
QVPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS
KADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

40 >ZV33_4 [<SEQ ID 922>] (SEQ ID NO: 922)
MKTFFKTLSSAALALILAACGGQKDSAPAASAAAPSADNGAEKKEIVFGTTVGDFGDMVK
EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEAF
QVPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS
KADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
45 SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

50 >ZV96_4 [<SEQ ID 923>] (SEQ ID NO: 923)
MKTFFKTLSSAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVQHKPYLDDFKKEHNLDITEVFQ
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDLDELGWIKLKDGINPLTASK
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEFGYKSPAAWNENGAAK*

Figure 8 shows the results of aligning the sequences of each of these strains. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. As is readily discernible, there is significant conservation among the

various strains of ORF 4 (SEQ ID NO: 216), further confirming its utility as an antigen for both vaccines and diagnostics.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

ABSTRACT

The invention provides proteins from *Neisseria meningitidis* (strains A & B) and from *Neisseria gonorrhoeae*, including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.

10

15

20

25

30

35